



DETECCIÓN DE NECESIDADES DE FORMACIÓN DE PROFESORES
EN EL USO DE UN SISTEMA DE GESTIÓN DEL APRENDIZAJE A
TRAVÉS DE MINERÍA DE DATOS

Pedro Ernesto Camacho Chacón

Tesis elaborada para obtener el grado de Maestro en Investigación Educativa

Tesis dirigida por
Alfredo Zapata González
Víctor Hugo Menéndez Domínguez

Mérida, Yucatán
Noviembre 2017

Oficio Aprobación del trabajo final

Mérida, Yucatán a 26 de mayo de 2017.

C. Pedro José Canto Herrera
Jefe de la Unidad de Posgrado e Investigación
Facultad de Educación, UADY
PRESENTE

Los abajo firmantes miembros del Comité Revisor nombrado por la dirección de la Facultad de Educación y en respuesta a su solicitud para revisar la tesis:

“Detección de necesidades de formación de profesores en el uso de un sistema de gestión del aprendizaje a través de minería de datos”

Presentado por Pedro Ernesto Camacho Chacón para obtener el grado de Maestro en Investigación Educativa, le comunicamos que el trabajo cumple con los requisitos de contenido y presentación establecidos por este Comité y por el Comité de Examen Profesional, de Especialización y de Grado, por lo tanto el dictamen que emitimos es de:


Aprobado

Por lo que puede proceder a la etapa de presentación y defensa del mismo.

Atentamente
Comité Revisor


Dr. Pedro José Canto Herrera
Miembro propietario


Dr. Víctor Hugo Menéndez Domínguez
Miembro propietario


Dr. Alfredo Zapata González
Asesor y Miembro propietario

C.c.p. Expediente del alumno en Control Escolar
C.c.p. Interesado

Declaro que esta tesis es mi propio trabajo, con excepción de las citas en las que he dado crédito a sus autores, asimismo afirmo que este trabajo no ha sido presentado para la obtención de algún título, grado académico o equivalente.

Pedro Ernesto Camacho Chacón

Agradezco el apoyo brindado por el Consejo Nacional de Ciencia y Tecnología (CONACYT) por haberme otorgado la beca No. 393709 durante el período de agosto de 2015 a julio de 2017 para la realización de mis estudios de maestría que concluye con esta tesis, como producto final de la Maestría en Investigación Educativa de la Universidad Autónoma de Yucatán.

Dedicatoria

Para mi familia, amigos, alumnos, y en especial a mi pareja, quien con su paciencia día a día consigue hacerme una mejor persona. A todos, ¡gracias! por su apoyo.

Agradecimientos

A mi asesor, el Dr. Alfredo Zapata González, por compartirme su tiempo, conocimientos y consejos, pero más aún por su trato profesional y al mismo tiempo empático.

A mi coasesor, el Dr. Víctor Hugo Menéndez Domínguez, por su invaluable tutoría en el pre procesamiento de los datos.

Al Departamento de Innovación e Investigación Educativa (DIIE) de la Universidad Autónoma de Yucatán (UADY), en especial a la M. en E.E. Jéssica B. Zumárraga Avila, por otorgarme el acceso a los datos que dieron origen al desarrollo del presente trabajo.

A la M.I.E. Maritza Minelli Briceño Caballero y a todos sus colaboradores, por la ayuda brindada para el mejoramiento de la información obtenida en el DIIE.

Al Mtro. Salvador Medina Peralta de la Facultad de Matemáticas de la UADY, por su dedicación y apoyo incondicional en los análisis inferenciales.

A los Dres. Manuel Emilio Prieto Méndez y José Ángel Olivas Varela de la Universidad de Castilla-La Mancha (UCLM), por permitirme vivir la experiencia de una estancia académica en el extranjero.

Y finalmente a la UADY, en particular a la Facultad de Educación y a todo el personal académico y administrativo que labora en la institución, por hacerme sentir en casa.

Resumen

Los Sistemas de Gestión de Aprendizaje (SGA) son una de las herramientas más empleadas por numerosas instituciones educativas de nivel superior para organizar y distribuir cursos en línea. Una problemática asociada a estos sistemas es que almacenan una gran cantidad de información de la interacción entre estudiantes y profesores que suele desaprovecharse.

Esta tesis está orientada al análisis del desempeño de 484 profesores de la Universidad Autónoma de Yucatán (UADY) durante el período enero - julio del año 2016, que utilizaron UADY Virtual como apoyo tecnológico a sus actividades docentes en los niveles de pregrado y posgrado.

La actividad de los profesores en UADY Virtual fue analizada a través del Método de Extracción de Conocimiento en Bases de Datos (KDD, por sus siglas en inglés), del cual se obtuvieron perfiles de profesores, reglas de asociación y clasificación.

Adicionalmente se realizó un análisis descriptivo para la obtención de indicadores estadísticos y un análisis inferencial para determinar si existieron diferencias significativas en el uso por parte de los profesores de UADY Virtual de acuerdo a los cinco campus que conforman a la Universidad.

En cuanto a los resultados obtenidos se observó que la gran mayoría de los profesores utilizaron UADY Virtual como un repositorio de tareas y materiales audiovisuales, a pesar de que algunos elementos destacaron de forma moderada, tales como los foros y cuestionarios.

Palabras clave: Sistemas de Gestión del Aprendizaje, MOODLE, Minería de datos.

Tabla de contenido

Dedicatoria / ix

Agradecimientos / xi

Resumen / xiii

Tablas / xvi

Figuras / xvii

Capítulo 1. Introducción / 1

1.1 Antecedentes / 1

1.2 Planteamiento del problema / 5

1.3 Pregunta de investigación, hipótesis y objetivos / 8

1.4 Justificación / 9

1.5 Limitaciones / 10

1.6 Método de trabajo / 10

1.7 Organización del documento / 11

Capítulo 2. Estado del arte / 12

2.1 Fundamentación pedagógica / 12

2.2 El aprendizaje electrónico y su evolución / 15

2.3 Sistemas de gestión del aprendizaje / 20

2.4 Extracción de conocimiento en bases de datos / 28

2.5 Minería de datos / 30

2.5.1 Técnicas de minería de datos / 31

2.5.2 Minería de datos en educación / 32

2.6 Investigaciones relacionadas / 36

Capítulo 3. Metodología / 41

3.1 Fase 1: Selección / 42

3.2 Fase 2: Preproceso / 44

3.3 Fase 3: Transformación / 51

3.4 Fase 4: Minería de datos / 52

WEKA / 52

SPSS / 54

3.4.1 Agrupamiento / 54

3.4.2 Clasificación / 55
3.4.3 Asociación / 55
3.5 Fase 5: Análisis de los resultados de las técnicas de minería de datos / 55
3.6 Fase 6: Análisis descriptivo e inferencial / 56
Capítulo 4. Resultados de la Minería de Datos / 57
4.1 Técnicas de agrupamiento / 57
4.2 Técnicas de clasificación / 62
4.3 Técnicas de asociación / 63
Capítulo 5. Resultados del Análisis Descriptivo e Inferencial / 65
5.1 Análisis descriptivo / 65
5.2 Análisis inferencial / 75
5.2.1 Herramientas con diferencias significativas de acuerdo al campus / 78
5.2.2 Diferencia del uso de UADY Virtual de acuerdo al campus / 80
Capítulo 6. Conclusiones / 81
6.1 Análisis de la consecución de los objetivos / 81
6.2 Discusión / 85
6.3 Recomendaciones y trabajo futuro / 86
6.4 Divulgación de resultados / 88
Referencias / 89

Tablas

Tabla 1. Comparación de las investigaciones relacionadas / 39

Tabla 2. Atributos de las bases de datos / 43

Tabla 3. Características de los grupos de profesores / 57

Tabla 4. Características de los grupos de profesores. Atributos con un uso mayor a 1 / 60

Tabla 5. Resultados obtenidos con el algoritmo de agrupamiento SimpleKmeans a través del software WEKA / 61

Tabla 6. Reglas de asociación de las herramientas de UADY Virtual / 63

Tabla 7. Actividades y recursos empleados al menos una vez por área de conocimiento / 68

Tabla 8. Porcentaje de uso de las actividades y recursos del sistema UADY Virtual / 69

Tabla 9. Prueba de Kolmogorov – Smirnov para cada una de las herramientas / 76

Tabla 10. Nomenclatura Áreas de Conocimiento (AC) / 78

Tabla 11. Total de herramientas usadas entre áreas de conocimiento / 80

Figuras

- Figura 1. Esquema general de la tesis / 11
- Figura 2. Evolución del aprendizaje en línea / 14
- Figura 3. Taxonomía del aprendizaje electrónico / 20
- Figura 4. Transformación de datos en información / 28
- Figura 5. Proceso para la extracción de conocimiento en bases de datos / 30
- Figura 6. Participantes en la Minería de datos educativos / 33
- Figura 7. Metodología KDD / 42
- Figura 8. Imagen parcial de las bases de datos que contenían los registros de acceso de los profesores en UADY Virtual / 43
- Figura 9. Vista parcial del uso de la función Subtotal obtenida con el software Excel / 48
- Figura 10. Vista parcial de la base de datos UADY Virtual V2.0 / 49
- Figura 11. Vista parcial de la base de datos UADY Virtual V2.0 / 50
- Figura 12. Base de datos UADY Virtual V4.0 / 51
- Figura 13. Uso del sistema UADY Virtual de acuerdo a los grupos de profesores / 58
- Figura 14. Uso del sistema UADY Virtual de acuerdo a los campus / 60
- Figura 15. Reglas de clasificación / 62
- Figura 16. Profesores que impartieron al menos un curso con apoyo del sistema UADY Virtual / 65
- Figura 17. Profesores habilitados por el Módulo 5 del MEFI / 66
- Figura 18. Porcentaje de uso de las actividades y recursos en UADY Virtual / 70
- Figura 19. Porcentaje de uso de las actividades y recursos en UADY Virtual dividido por áreas de conocimiento / 71
- Figura 20. Porcentaje de las actividades de uso alto / 72
- Figura 21. Porcentaje de las actividades de uso bajo / 72
- Figura 22. Porcentaje de los recursos de uso alto / 73
- Figura 23. Porcentaje de los recursos de uso bajo / 73
- Figura 24. Distribución de uso de las herramientas de UADY Virtual para cada área de conocimiento / 74
- Figura 25. Prueba de normalidad para la variable global (total de herramientas usadas) / 76

Capítulo 1. Introducción

El presente capítulo ofrece una descripción general de la tesis. Se expondrá el antecedente que dio motivos para el planteamiento de la problemática y derivado de ello, su pregunta de investigación; asimismo se describirá la hipótesis, los objetivos de investigación y el método de trabajo empleado. Finalmente se presentará visualmente la estructura de la organización del documento.

1.1 Antecedentes

A continuación se exponen los precedentes de la investigación con la finalidad de contextualizar la problemática desde diversos panoramas, el social, político, científico y disciplinar, tanto a nivel global como regional.

En las últimas décadas, el tema de la comunicación en la sociedad se ha transformado aceleradamente debido a los avances tecnológicos y al acceso (cada vez más fácil) que tienen las personas para usar tanto la Internet como los dispositivos digitales.

Drucker (1992), afirma que el concepto de sociedad del conocimiento surge a consecuencia de la implementación de las Tecnologías de Información y Comunicación (TIC) en la vida cotidiana, y afecta, las relaciones sociales, culturales y económicas, en donde se eliminan las barreras de espacio y tiempo, y se facilita una comunicación ubicua y asíncrona.

De acuerdo con Hilbert (2012), se puede hablar de cinco olas en la evolución de las tecnologías actuales. La primera impulsada por la revolución industrial, la segunda por la invención de la máquina de vapor, la tercera por “la electrificación de las organizaciones sociales y productivas”, la cuarta por la “motorización y la movilización automática de la

sociedad” y la quinta por la “digitalización de los procesos de información y comunicaciones en los sistemas sociales” (p. 262).

Por su parte, las instituciones educativas de diferentes países se han ajustado desde principios de los años noventa a dicha influencia impulsada por la expansión de las TIC. La globalización ha tendido y promovido una nueva cultura mundial, en donde se rompen las barreras geográficas y económicas (García, 2013).

La Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), expone que las TIC contribuyen a la mejora de la educación en el mundo, tanto al facilitar el acceso universal de la misma y el desarrollo de los profesores, como en la eficiencia en los aspectos administrativos y de gestión escolar. De igual manera, la UNESCO promueve que la educación en línea es pieza clave para lograr una sociedad basada en el conocimiento (UNESCO, 2002).

A la par, se pueden apreciar políticas de desarrollo humano que promueven el derecho al acceso a Internet y al uso de las TIC desde la infancia, para que los niños se incorporen íntegramente a la nueva sociedad del conocimiento y de la información (CEPAL. División de Desarrollo Social, 2014).

Particularmente a mediados de los años noventa, en los Estados Unidos de América (EUA), se comenzaron a utilizar recursos económicos para incrementar el acceso a Internet, y a finales de la década, más del 90% de las escuelas contaban con computadoras y conexiones a ésta (Anderson & Becker, 2001).

La implementación de las TIC en un contexto educativo puede tener efectos positivos; por ejemplo, una investigación desarrollada por Kulik (1994), demostró que los

estudiantes que utilizaron computadoras en clase obtuvieron puntajes más altos en pruebas de aprovechamiento y tuvieron una percepción positiva de sus cursos.

De igual manera en los EUA, Schacter (1999) y Wenglinsky (1998), capacitaron a docentes en el uso de TIC para su implementación en el aula, lo que trajo como resultado que sus estudiantes lograran mayores niveles de logro en comparación con los docentes que no usaron la tecnología.

En América Latina la educación apoyada de las TIC también ha sido un motivo de estudio. En la década de los años 90 se abrió paso a las políticas públicas que vinculan las TIC con la educación. En los últimos años, la región ha sido muy dinámica en la integración de las tecnologías a su sector educativo, por lo que han surgido políticas que se establecen principalmente en tres experiencias: 1. Reducción de la brecha digital, 2. Democratización del acceso a las TIC y 3. La consideración de la situación en la que se encuentre cada país al momento de implementar las iniciativas (Lugo, López y Toranzo, 2014).

Asimismo, el Banco Mundial (2013) comienza a apoyar la introducción de las TIC en el aula para los países de América Latina, con el objetivo de expandir el acceso a la educación, para en consecuencia, abatir la pobreza al lograr incrementar las oportunidades de empleo de los ciudadanos, al prepararlos con los conocimientos de alcance mundial más recientes.

En el plano Nacional, se pueden apreciar políticas en México encaminadas a incorporar las TIC en el desarrollo del sector educativo, tales como el Programa Especial de Ciencia, Tecnología e Innovación (2014-2018). Este programa, respaldado por el Plan

Nacional de Desarrollo 2013-2018, tiene como punto de partida el objetivo de “hacer del desarrollo científico, tecnológico y la innovación, pilares para el progreso económico y social sostenible” (p.5), con lo que se persigue una educación de calidad. Para ello el gobierno federal estableció cinco estrategias, entre ellas “contribuir al fortalecimiento de la infraestructura científica y tecnológica del país (p.41). De lo anterior radica la importancia que México le está dando a la incorporación de las TIC a la educación del país (Programa Especial de Ciencia, Tecnología e Innovación 2014-2018).

A nivel local, la Universidad Autónoma de Yucatán (UADY), ha implementado políticas para el uso de las TIC en sus procesos académicos y administrativos, las cuales se contemplan en su Plan de Desarrollo Institucional (PDI) 2010-2020, 2014-2022 y en su Modelo Educativo para la Formación Integral (MEFI).

En el PDI del 2010-2020 (UADY, 2010), la universidad plantea lo siguiente: “... la formulación de nuevos programas de licenciatura y posgrado utilizando modalidades no presenciales y semipresenciales, privilegiando la equidad y el uso de tecnologías de información y comunicación. La nueva oferta debe considerar su contribución a la operación de los programas presenciales, así como a la formación de los estudiantes inscritos en ellos” (p.76).

Por su parte, el MEFI (UADY, 2012) señala: “... la planeación deliberada y sistemática de nuevas propuestas para la solución de situaciones problemáticas y para la mejora continua de la práctica educativa que implica un cambio en el contexto y la práctica educativa misma, mediante la incorporación de recursos y medios educativos vanguardistas...” (p.39); respecto al profesor universitario: “Utilizar las tecnologías de

información y comunicación como recurso didáctico en su ejercicio docente de manera pertinente.” (p.49).

En respuesta a lo anterior, la UADY integró la educación virtual como una herramienta de apoyo tecnológico a las clases presenciales en los niveles de pregrado y posgrado, esto dio origen a un entorno virtual de aprendizaje denominado UADY Virtual, el cual está basado del software de distribución libre Moodle. Esta herramienta tecnológica en conjunto con el MEFI, plantean el siguiente objetivo: “Impulsar la innovación académica de la institución en todos los niveles, a través de las diferentes modalidades educativas contempladas en el MEFI, especialmente con el apoyo de entornos virtuales de aprendizaje y la puesta en marcha de las modalidades mixta y no presencial (modalidades no convencionales).” (UADY, 2013, p.31).

Como se ha señalado, la UADY, de acuerdo a las políticas locales, nacionales e internacionales, contempla la implementación de las TIC para sus procesos de enseñanza y aprendizaje, así como también ampliar su oferta educativa a través de los entornos virtuales de aprendizaje. En este sentido, esta tesis se enfoca en un tema fundamental para dicho contexto, como lo es, la identificación de necesidades de formación, que los profesores pudiesen requerir en la implementación de TIC en sus aulas.

1.2 Planteamiento del problema

De acuerdo con el Plan Desarrollo de UADY Virtual, la Universidad ha desarrollado diversas acciones en términos de formación de capital humano, a continuación se mencionan aquellas acciones que han impactado en la formación de los docentes respecto al uso de las TIC y de las plataformas educativas (UADY, 2013).

En el año 2003, la Dirección General de Desarrollo Académico (DGDA) impartió cursos para la habilitación de profesores para el uso del Sistema de Gestión del Aprendizaje (SGA), en ese entonces llamado Dokeos, "... así como en el diseño instruccional para la enseñanza en entornos virtuales." (p.24).

En el año 2007 se desarrolla e implementa el Módulo V, conocido como: "Innovación educativa: Uso de las TIC en educación" el cual "... ha formado a 145 profesores de los cinco campus y del Centro de Investigaciones Regionales Dr. Hideyo Noguchi, para el uso de la plataforma Dokeos o MOODLE, bajo diferentes diseños instruccionales." (p.24).

Respecto a las tutorías en línea, entre los años 2009 y 2010 se capacitaron a 96 profesores a través del curso "*e tutoring*" impartido por el Consejo Británico; de igual manera, en ese mismo periodo se habilitaron a profesores de las Facultades de Economía, Odontología, Enfermería y Ciencias Antropológicas para el uso de la plataforma Moodle.

En el año 2011, la UADY estableció un convenio con Virtual Educa Cono Sur para especializar a un grupo de académicos y administrativos de TI en entornos virtuales de aprendizaje, dicha capacitación se realizó a través del Instituto Argentino de Formación Docente.

En relación con las necesidades de formación de profesores para el desarrollo de competencias en diseño instruccional, tutoría y tecnológicas para las modalidades no convencionales, es decir, no presencial y mixta, la UADY condujo una investigación para la cual se encuestó a 924 académicos de 15 Facultades, dos escuelas preparatorias y la unidad académica de bachillerato con interacción comunitaria.

Los principales resultados, de acuerdo con el Plan Desarrollo de UADY Virtual (UADY, 2013), fueron los siguientes:

1. El 40% del personal académico consideró tener las competencias para poder desarrollar un curso en una modalidad no convencional.
2. El 38% indicó poder desarrollarse en un nivel elemental como tutor en línea.
3. El 80% manifestó poder hacer uso, al menos en un nivel elemental, de los recursos tecnológicos necesarios para la modalidad no convencional.
4. El 80% aseguró estar interesados en participar en las modalidades no convencionales.
5. El personal académico, en los comentarios, recomendó mayor capacitación, reorganización de las funciones y mejores servicio tecnológicos.

Como se ha observado, la UADY, consciente de la importancia del uso de las TIC y particularmente, de los SGA, ha desarrollado diversas acciones en función a la detección de necesidades docentes, y en consecuencia la capacitación de estos, sin embargo, a pesar de los esfuerzos antes descritos, existe un área de oportunidad que consiste en el análisis de la actividad realizada por los profesores en el desarrollo de sus cursos en línea, los cuales en su gran mayoría se utilizan como una herramienta tecnológica de apoyo de los cursos presenciales.

1.3 Pregunta de investigación, hipótesis y objetivos

A continuación se presenta la pregunta de investigación relacionada con la problemática planteada, asimismo la hipótesis en función a la pregunta de investigación, y los objetivos, tanto los específicos como el general.

Pregunta de investigación

¿Existen necesidades de formación que pudiesen estar requiriendo los profesores de la UADY de acuerdo a la utilización del sistema de gestión de aprendizaje UADY Virtual?

Con base a la pregunta de investigación se planteó la siguiente hipótesis:

Hipótesis

Los profesores de la Universidad Autónoma de Yucatán requieren formación en el uso las herramientas del sistema de gestión de aprendizaje UADY Virtual.

A partir de la hipótesis establecida se establecieron tanto el objetivo general de la tesis así como los objetivos específicos:

Objetivo general

Determinar si existen necesidades de formación en el uso de las herramientas del sistema de gestión de aprendizaje UADY Virtual.

Los objetivos específicos de la investigación fueron:

1. Crear perfiles de profesores de acuerdo a su desempeño con el sistema UADY Virtual.

2. Construir un modelo de conocimiento que permita predecir el desempeño del profesor en función de la cantidad de veces que emplea ciertas actividades y/o recursos del sistema UADY Virtual.
3. Construir reglas de asociación de las relaciones entre las herramientas de UADY Virtual.
4. Calcular las frecuencias y porcentajes de las actividades y recursos usados por los profesores de la UADY y presentarlas de acuerdo a las cinco áreas de conocimiento.
5. Determinar si existen diferencias significativas en el uso de las actividades y recursos del sistema por parte de los profesores de la UADY en función de las cinco áreas de conocimiento.

1.4 Justificación

Generalmente los Sistemas de Gestión del Aprendizaje (SGA) son fáciles de usar, se consideran flexibles en términos pedagógicos y eficientes con relación a los costos, además permiten organizar los contenidos educativos, a la vez que ofrecen una serie de funcionalidades de comunicación (Downes, 2005).

Los SGA resguardan grandes cantidades de información relacionada con los profesores y estudiantes, por ejemplo, horarios de conexión, cantidad de páginas visitadas, duración de estadía en cada página, tipos de herramientas empleadas, número de tareas entregadas, mensajes compartidos en foros, entre otros.

De acuerdo con Trandafilu (2012), las Instituciones de Educación Superior (IES) se encuentran abrumadas por la gran cantidad de información almacenada en sus entornos

virtuales de aprendizaje. Lo anterior ha originado una complejidad en el proceso de análisis de información. Transformar en conocimiento la actividad que se almacena cada día dentro del entorno de los sistemas representa un gran desafío científico y tecnológico.

Desafortunadamente, en muchas ocasiones las instituciones educativas no aprovechan del todo la información acumulada en los sistemas. Por lo tanto, es necesario realizar investigaciones que incorporen como parte de su proceso metodológico, extraer conocimiento de la información almacenada en los SGA.

La presente investigación considera que la cantidad del uso de las herramientas de UADY Virtual por parte de los profesores representa el dominio que tienen sobre dichas herramientas, sin embargo, se reconoce que el no uso de dichas herramientas no necesariamente implica falta de competencia del profesor, pero por otro lado, evidencia su desinterés, y por tal la pertinencia el desarrollo de la presente tesis, ya que se desea explorar el uso que los profesores le dan a UADY Virtual.

1.5 Limitaciones

La investigación no tiene como propósito identificar las causas que pudiesen explicar el grado de utilización de UADY Virtual por parte de los profesores. Se parte de un escenario ideal en donde la tecnología en sí misma no representa una limitación, particularmente la conectividad a internet.

1.6 Método de trabajo

De manera general se enlistan a continuación las principales fases consideradas para la realización de la tesis:

1. Definición de la problemática.

2. Análisis del estado de arte.
3. Diseño metodológico de la investigación.
4. Experimentación.
5. Difusión.

1.7 Organización del documento

Considerando el método de trabajo expuesto previamente, la tesis ha sido estructurada de la siguiente manera (Fig. 1):

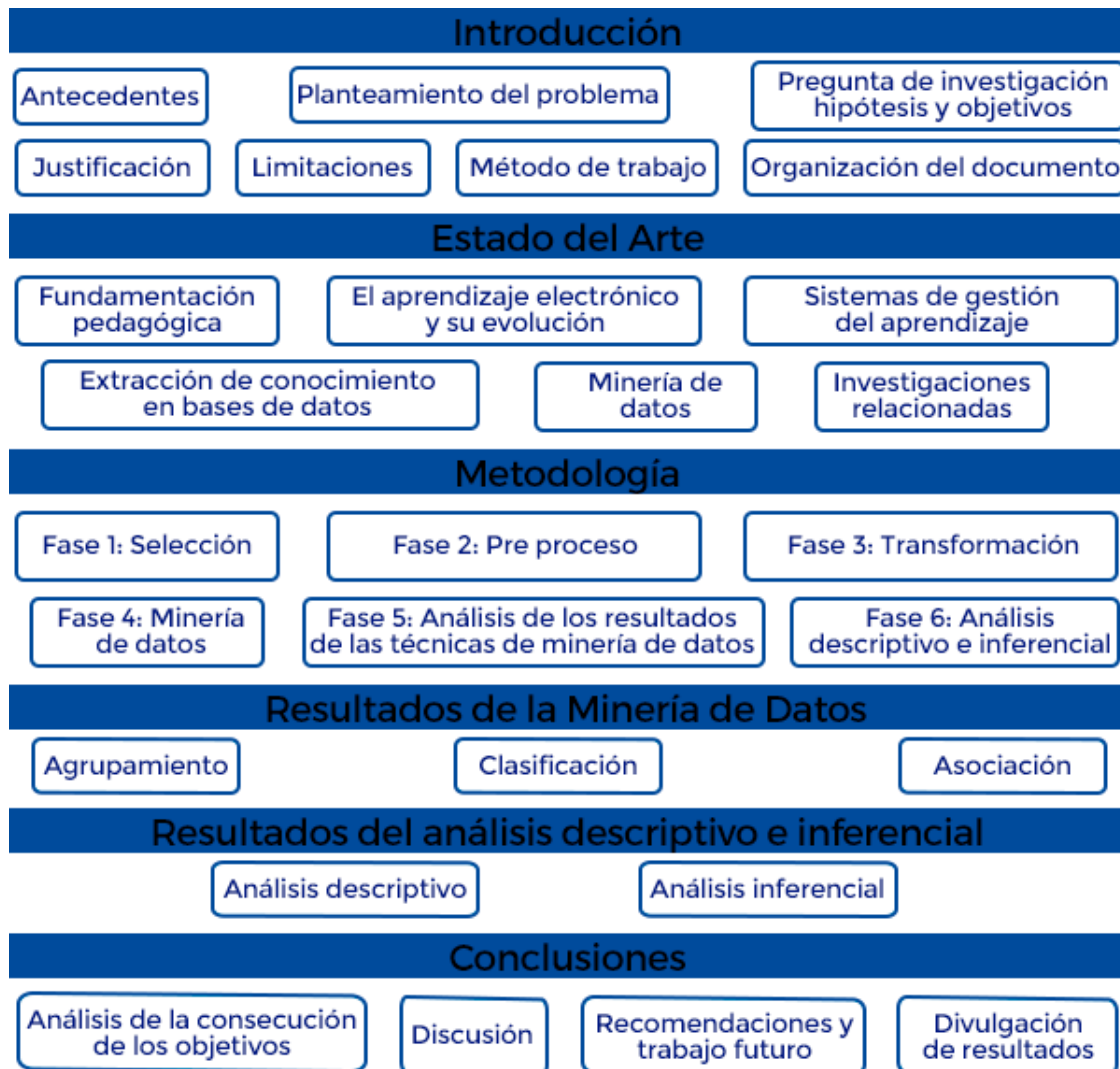


Figura 1. Esquema general de la tesis.

Capítulo 2. Estado del arte

El objetivo de este capítulo es presentar los conceptos, definiciones y trabajos relacionados con el tema de investigación. Se inicia con una introducción, donde se expone el aprendizaje en línea, se describen sus conceptos relacionados, historia y fundamentación pedagógica. Posteriormente, se describen las características de los Sistemas de Gestión del Aprendizaje, así como algunos ejemplos de dichos sistemas; enseguida, se explica el proceso para la extracción de conocimiento en bases de datos, para el cual se profundizará en la técnica de minería de datos. Finalmente, se exponen los trabajos relacionados a la investigación.

2.1 Fundamentación pedagógica

De acuerdo con Marenzi (2014), la educación tecnológica se remonta al año 1954, fecha en la que el psicólogo Skinner publica: “*The science of learning and the art of teaching*”. Una de sus principales premisas radica en que el comportamiento no debe descansar en la hipótesis del estado interno del individuo, sino en la observación descriptiva de sus conductas; por consiguiente, es posible programar el aprendizaje del estudiante, el cual requiere desarrollar actividades de acuerdo a su propio ritmo. Dicha corriente psicológica, basada en el comportamiento y su programación gracias a reforzadores, es conocida como conductismo.

En respuesta a lo anterior, el cognoscitivismo surge a finales de los años cincuenta, dicha teoría sitúa el aprendizaje humano analógicamente a la manera en la que las computadoras procesan información. Se desea el entendimiento de los procesos meta-cognitivos, así como de estrategias para su fortalecimiento, por ejemplo, organizadores

gráficos. El principal representante de dicha corriente fue el psicólogo Jerome Bruner (Marenzi, 2014).

Al inicio de los noventas, el constructivismo surge como un nuevo modelo para la explicación del aprendizaje humano, basa sus ideas en el supuesto de que la adquisición de conocimiento debe ser un proceso de construcción activa del mismo, por lo tanto, la instrucción del profesor debe despertar interés por parte del estudiante para ser descubierto por él mismo, a diferencia de ser expuesto por el primero (Marenzi, 2014).

Siemens (2004), propone un nuevo enfoque de aprendizaje, el “conectivismo”, que se caracteriza por la amplificación del aprendizaje, conocimiento y entendimiento, a través del desarrollo de redes personales. En el conectivismo el aprendizaje del estudiante deja de ser únicamente interno, se reduce la dependencia del involucramiento del profesor y se valora la red de nodos creados por el estudiante.

De acuerdo a la evolución de las teorías de aprendizaje descritas anteriormente, son notorios varios aspectos, en primer lugar y en cuanto a la didáctica docente, ésta ha pasado de ser discursiva a directiva, es decir, el profesor (bajo los modelos pedagógicos actuales, por ejemplo, competencias) centra su didáctica en el desarrollo de actividades de aprendizaje, y deja a un lado las estrategias educativas tradicionales, tales como el discurso, dictado, entre otras.

En segundo lugar, el contenido ha pasado de ser aplicable a todos los estudiantes, a un contenido específico de acuerdo a las necesidades e intereses particulares de los alumnos. Por otro lado, la evaluación, actualmente se centra en la observación de destrezas y productos entregables por el estudiante, por ejemplo, a través de la estrategia denominada

portafolio de evidencias, en comparación con los métodos tradicionales de evaluación tales como las pruebas escritas.

Finalmente, respecto a los entornos tradicionales de aprendizaje y participación entre estudiantes, éstos han evolucionado hacia entornos de aprendizaje basados en comunidades colaborativas a través de redes sociales y formación de grupos de estudio.



Figura 2. Evolución del aprendizaje en línea.

Para el desarrollo de lo anterior, las tecnologías educativas digitales basadas en Internet, cobran un sentido de necesidad, ya que permiten la ejecución de las interacciones previamente señaladas, facilitadas por herramientas sincrónicas y/o asincrónicas. A continuación se detallan las características de dichas tecnologías, y posteriormente las herramientas que permiten el análisis de su información para un posterior uso educativo.

2.2 El aprendizaje electrónico y su evolución

De acuerdo con Cooke (2004), el aprendizaje electrónico (en inglés, *e-learning*) surge como necesidad por parte de las empresas para capacitar a sus trabajadores. Alrededor de los años noventa, diversas compañías comenzaron a usar videos para sus cursos de capacitación, sin embargo, presentaban varias limitaciones: 1) incapacidad por ajustar el contenido a las necesidades de los trabajadores, 2) costo de mantenimiento elevado y 3) el contenido no podía ser actualizado fácilmente. Adicionalmente se requería del equipo adecuado para ello, y el seguimiento y evaluación de los trabajadores resultaba difícil (Hoffmaster, 2006).

En respuesta a las limitaciones de los videos, surgió el *Computer Based Training* (CBT), con tecnologías tales como: Windows 3.1, Macintosh, discos compactos y paquetería informática, que marcaron el origen del desarrollo de la era de la multimedia (Kiffmeyer, 2004).

Los discos compactos redujeron los costos, los empleados podían completar su entrenamiento al usar dicho recurso en sus computadoras personales; sin embargo, aún era imposible rastrear el avance del aprendizaje del trabajador. Lo anterior comenzó a resolverse gracias al uso de la Internet, al emplearlo como medio para el almacenamiento de información, sin embargo, en un comienzo, la interfaz era muy simple, únicamente permitía texto e imágenes de poco peso en bits (Cooke, 2004).

Con el continuo avance de las TIC y con la introducción de estrategias pedagógicas a los cursos de capacitación, cuyo contenido se distribuía a través de la Internet, el

aprendizaje electrónico evoluciona, y se denomina como aprendizaje en línea (en inglés, *online learning*) (Clark, 2002).

En la literatura se encuentran diversas definiciones sobre aprendizaje en línea; ante esta diversidad se han seleccionado algunas de ellas:

1. Es el uso de nuevas tecnologías de multimedia e Internet para el mejoramiento de la calidad del aprendizaje al facilitar el acceso a recursos y servicios, así como el intercambio y la colaboración (Alonso, 2005).
2. Es el proceso educacional que utiliza a las TIC para mediar el aprendizaje a través de actividades síncronas y asíncronas (Jereb, Smitek, 2006).
3. Es definido como las TIC usadas para el soporte del estudiante, con el propósito de mejorar su aprendizaje (Ellis, Ginns y Piggot, 2009).

De acuerdo con Sangra, Vlachopoulos & Cabrera (2012), la definición del aprendizaje en línea posee cuatro dimensiones para su interpretación, a saber: enfoque tecnológico, de sistemas, de comunicación y educativo.

En la actualidad existen diversos términos relacionados con el aprendizaje en línea, al mismo tiempo que diversos autores emplean de múltiples maneras dichos conceptos, lo que puede causar confusión respecto a la terminología; en consecuencia, a continuación se describen los principales términos relacionados con el concepto de aprendizaje en línea y se adopta una postura en cuanto al empleo del vocabulario para la presente investigación.

El autor Naidu (2006), separa el concepto de aprendizaje electrónico de aprendizaje en línea, ya que asocia al primero con únicamente tecnologías informáticas, mientras que al segundo con el medio para su distribución (Internet). Al respecto, para la presente

investigación se considerarán sinónimos los conceptos aprendizaje electrónico (del inglés, *e-learning*) y aprendizaje en línea, debido al uso de la Internet como medio indispensable para el desarrollo del aprendizaje digital.

Por otro lado el aprendizaje en línea se relaciona, de acuerdo con Dillenbourg (2000), al aprendizaje virtual (en inglés, *virtual learning*), el cual a su vez se relaciona con ambientes virtuales de aprendizaje (en inglés, *virtual learning environments*), y se refiere a los espacios diseñados para el consumo de información virtual, los cuales deben ser sociales y motivadores para el aprendizaje, cuyos estudiantes sean partícipes de su propio conocimiento. Cabe destacar que dicho concepto no está limitado a la educación en línea, ya que se asocia a las clases presenciales con apoyo de medios informáticos.

Asimismo, en concordancia con Lynch (2002), el aprendizaje en línea se relaciona con la educación basada de Internet (en inglés, *web based education*); misma que es totalmente en línea, ya que no existe interacción física entre el estudiante y profesor.

De igual manera, surgen otros conceptos relacionados en función a otras dimensiones, por ejemplo, de acuerdo al dispositivo empleado. En éste sentido, el aprendizaje en línea se relaciona con el aprendizaje móvil (en inglés, *mobile learning, m-learning*), el cual se refiere a la educación a través del uso de dispositivos móviles, tales como teléfonos inteligentes y tabletas.

De acuerdo a su impacto en la modalidad de la clase, el aprendizaje en línea se relaciona con los siguientes conceptos:

1. Aprendizaje combinado (en inglés, *blended learning*, *b-learning*): Combina las interacciones cara a cara entre profesor y estudiante, y el uso de interacciones en línea para el mismo curso (Shibley, 2009).
2. Clases invertidas (en inglés, *flipped learning environment*): Modelo educativo en donde el estudiante desarrolla las actividades de la casa en la escuela y viceversa (Honeycutt, 2013); es decir, las tareas que comúnmente se desarrollarían en la aula, se efectúan en el hogar. Lo anterior se logra con el apoyo de las TIC.

Desde el punto de vista tecnológico, el aprendizaje en línea se relaciona con los siguientes conceptos:

3. Sistemas de Gestión del Contenido (SGC, en inglés, *Content Management System*, CMS): Consiste en un sistema que colecta, almacena y publica contenido; asimismo ofrece diversas funcionalidades para las diferentes categorías de usuarios (Boiko, 2004).
4. Repositorios de objetos de aprendizaje (ROA): "Cualquier colección de recursos que son accesibles a través de una red sin el conocimiento previo de la estructura de la colección". (IMS Global Learning Consortium, 1997).
5. Sistema de Gestión del Aprendizaje (SGA, en inglés, *Learning Management System*, LMS): Sistemas que permiten organizar y distribuir materiales de un curso educativo, desarrollar foros de discusión, realizar tutorías, seguimiento y evaluación a los alumnos (Ortiz, 2007).

6. Sistemas de Gestión de Contenidos de Aprendizajes (SGCA, en inglés, *Learning Content Management System, LCMS*): “Son sistemas independientes o integrados con un LMS, que gestionan y administran los contenidos de aprendizaje. Una vez que los contenidos están en este sistema, pueden ser combinados, asignados a distintos cursos o descargados, entre muchas otras opciones” (Zapata, 2013).

7. *Adaptive Intelligent Learning System (AILS)*: Ya que se basan del aprendizaje en línea, proveen una experiencia personal de aprendizaje para cada estudiante de acuerdo a sus hábitos, los AILS pueden comunicarse entre sí, están integrados a los SGA y asociados con Objetos de Aprendizaje (OA) (Cemile, 2008).

A manera de conclusión, a continuación se presenta un esquema que permite identificar las relaciones entre los términos antes mencionados (figura 3).

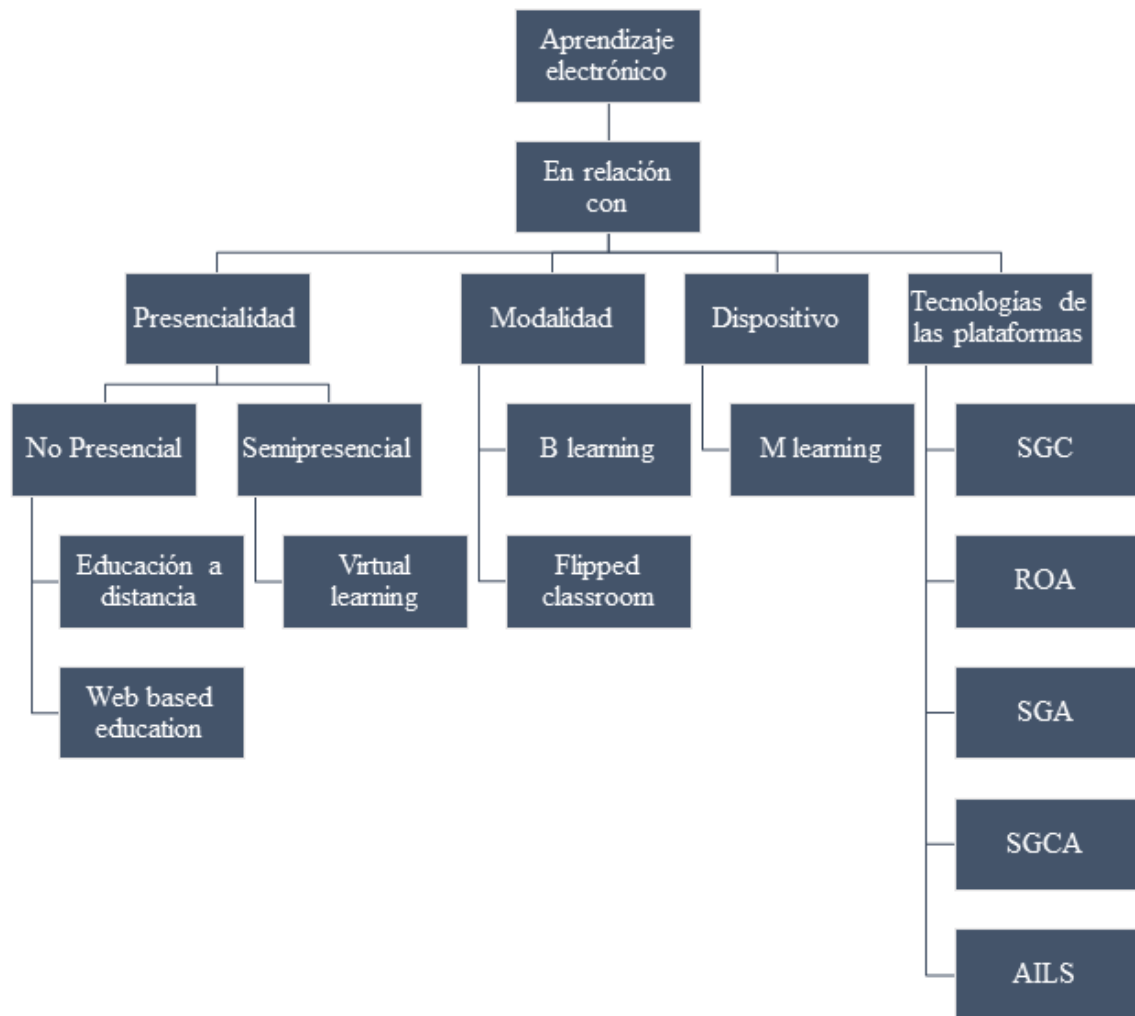


Figura 3. Taxonomía del aprendizaje electrónico.

Debido a que en la investigación se trabajará con un entorno educativo digital, basado en un sistema SGA, en la siguiente sección se describirán las características del sistema y algunos ejemplos estos.

2.3 Sistemas de gestión del aprendizaje

Abdullateef (2015), señala que un Sistema de Gestión de Aprendizaje (SGA, en inglés, *Learning Management System*, LMS) consiste en un sistema de comunicación que permite el desarrollo de un aprendizaje en línea. Su función principal es el almacenamiento

y distribución de material educativo, que soporta la administración y comunicación de la enseñanza y el aprendizaje. Los SGA son las herramientas tecnológicas más empleadas actualmente por numerosas instituciones para organizar y distribuir cursos en línea (Kapp, 2003).

Existen varias maneras de alojar un SGA y que este funcione, las principales son: el uso de un sistema complejo elaborado dentro de una institución, el cual estaría diseñado para llevar a cabo las funciones de gestionar cursos y el aprendizaje de los alumnos. De acuerdo con Sánchez (2009), los SGA contienen una gran variedad de herramientas, tales como:

1. Distribución de contenido: es un espacio en donde el profesor comparte contenidos, por ejemplo: documentos, imágenes, videos, enlaces a sitios de interés, entre otros.
2. Comunicación y colaboración síncrona y asíncrona: en las herramientas asíncronas el estudiante controla el momento para el consumo del contenido, mientras que en las síncronas se requiere conexión a la plataforma en un momento determinado, y la utilización de los contenidos es un por un espacio de tiempo pre establecido. El objetivo de las anteriores, permite que los participantes puedan comunicarse y trabajar colaborativamente. Algunos ejemplos son: foros, salas de chat, mensajería interna, wikis, diarios, etc.
3. Seguimiento y evaluación: en este rubro destacan las pruebas generadas por el profesor y autoevaluaciones del alumno.

4. Administración y asignación de permisos: posibilitan la inscripción, y por consiguiente, el registro al curso.
5. Complementarias: contiene diversas ayudas tales como el portafolio, el bloc de notas, los sistemas de búsqueda de contenidos, entre otras.

Adicionalmente a las características anteriores, Cabrero (2007) añade que son espacios de documentación compartida, debido a que posibilitan la administración de recursos por parte de profesores y estudiantes, por ejemplo, almacenamiento de documentos, y creación de carpetas para la gestión del contenido, entre otros. Para complementar la información anterior, las características claves de los SGA son (Ahmad, et al., 2005):

- Administración de varios cursos o programas.
- Llevar registro de los participantes.
- Organizar y administrar el aprendizaje de los participantes.
- Capacidad de proporcionar informes.
- Programar y organizar los tiempos para el aprendizaje.

Los SGA de acuerdo con su comercialización se clasifican en software propietario y software de distribución libre. Los primeros, se caracterizan por que han sido desarrollados por una empresa particular. Su modelo de negocio se basa a través de la licencia de uso se paga bajo un esquema mensual o anual (el costo depende del número de usuarios). Algunas de las ventajas son:

- Suelen ser más estables y con funcionalidades que pueden adaptarse de acuerdo a las necesidades y el presupuesto.

- Incluyen en muchos casos el alojamiento (servidores y ancho de banda) desligando al cliente de lidiar con estos requerimientos.
- La capacitación de los profesores corre a cargo de la empresa contratada.

Algunas de sus desventajas son:

- Suelen ser muy costosas: hay empresas que no solo cobran un mantenimiento mensual por el alojamiento en sus servidores sino también un costo mínimo de implementación.
- No se tiene acceso al código fuente ya que las empresas no suelen vender el software, sino su implementación y posterior administración y mantenimiento, con el fin de asegurarse un abono mensual y/o anual.

Algunos ejemplos de Sistemas de Gestión del Aprendizaje de software propietario son:

Blackboard, es la plataforma de código privado líder en el mercado. Este SGA se caracteriza por poseer tecnologías educativas orientadas a una interfaz social, que permita la participación activa y comprometida de sus participantes. Asimismo permite analíticas y herramientas de comunicación y soporte. Este sistema tiene presencia en más de 2200 instituciones educativas en más de 60 países. En México, diversas instituciones privadas lo utilizan, tales como el Tecnológico de Monterrey, Universidad Iberoamericana, Universidad Anahuac, entre otras (Blackboard, 2016).

Dokeos, en sus inicios fue un SGA de código abierto, sin embargo hace algunos años cambio su filosofía y se convirtió en un software cerrado. Este sistema posee una interfaz intuitiva que facilita la creación y organización de contenidos interactivos y ejercicios, así como también, incorpora herramientas de seguimiento e informes sobre el

desempeño de los alumnos en el curso. Su arquitectura está basada en módulos lo que facilita la incorporación de nuevas funciones (De Praetere, 2010).

E-ducativa, es una empresa de tecnología informática orientada a la implementación de soluciones de e-learning desde hace más de 10 años. Este sistema está orientado al mercado Hispano. Actualmente, cuenta con soluciones tanto para instituciones educativas como para empresas y gobiernos (educativa, 2016).

Sumtotal (Skillsoft Company, 2016), es un SGA con un alto grado de implantación en el ámbito empresarial. Incluye la posibilidad de integrarlo con la gestión de recursos humanos de la empresa. También cuenta con adaptación a dispositivos móviles.

eCollege (Pearson education Inc, 2016), este SGA brinda a docentes y administradores las herramientas, el contenido y el apoyo que necesitan para crear, gestionar y medir experiencias de aprendizaje personalizadas y atractivas para sus estudiantes. Este sistema incluye herramientas para el apoyo del análisis de la gestión de programas educativos y ofrece seguimiento personalizado a los objetivos de aprendizaje.

Los SGA cuyo software es de distribución libre pueden instalarse tanto en computadoras personales como en servidores web. El soporte técnico suele estar soportado por una comunidad mundial de desarrolladores. Algunas de sus ventajas son:

- Permite realizar modificaciones sobre el código del sistema, ya sea para personalizarlo o agregarle nuevos componentes.
- Son compatibles con la mayoría de los formatos estándares.
- Su actualización es constante.

- Ausencia de malware al momento de instalarlo.

En contraste, algunas de sus desventajas son:

- Carece de soporte técnico, en caso de que se presente algún problema, se puede consultar en foros especializados en la web.
- Requiere implementar una infraestructura por parte de la institución que lo adopta.
- La capacitación de los profesores corre a cargo de la institución que adopta el sistema.

Algunos ejemplos de Sistemas de Gestión del Aprendizaje de distribución libre o código abierto son:

ILIAS (Integrated Learning, Information and cooperAtion System) (University of Cologne, 2004), es un sistema de gestión del aprendizaje de código abierto. Posee un ambiente de aprendizaje flexible, el cual permite la reutilización y el intercambio de materiales de instrucción entre docentes y alumnos. Destaca su sistema de control de acceso basado en roles y la facilidad para la creación de cuestionarios.

ATutor (Toronto University, 2002), es un sistema de código abierto que se utiliza para la impartición de cursos en línea. Su instalación es rápida y pueden desarrollarse temas para cambiar la apariencia de la plataforma así como instalar nuevos módulos para ampliar las funcionalidades de la misma. Los docentes pueden crear, configurar y empaquetar cursos para su redistribución en la red y los estudiantes aprenden en un entorno adaptativo y social.

Sakai (Apereo Foundation, 2016), Consiste en un sistema que permite la enseñanza, aprendizaje, investigación y colaboración educativa; posee una interfaz simple, la cual puede ser personalizada por cada institución, asimismo cuenta con múltiples funciones tales como: tareas, calendarios, chat, foro de discusión, lecciones, pruebas, entre otras.

MOODLE (*Modular object oriented distance learning environment*) (Dougiamas y Taylor, 2003), es un sistema de gestión de cursos de aprendizaje en línea de distribución libre. Su diseño está basado del enfoque pedagógico constructivista social. Posee una arquitectura modular, lo que permite incorporar una gran diversidad de componentes y funcionalidades. Implementa una interfaz de navegador de tecnología sencilla, ligera, eficiente y compatible.

La fase experimental de este trabajo de investigación se desarrollará en el entorno de un Sistema de Gestión del Aprendizaje basado en el software MOODLE.

Los SGA del futuro combinarán cinco tendencias basadas en (Graduarte XXI, 2016):

- Las escuelas híbridas.
- Las plataformas adaptativas.
- Las estrategias de gamificación y nuevos certificados.
- Las redes sociales y la creación colaborativa de aprendizajes.
- La corrección automática de evaluaciones.

Algunos ejemplos de estos Sistemas de Gestión del Aprendizaje son:

Fishtree, es un sistema que además de tener todas las funciones de los clásicos gestores de aprendizaje, tiene características que permiten al docente diseñar planes de

clases con base en distintos recursos digitales, esto permite que se genere nuevas formas de impartir una clase y promueva el aprendizaje adaptativo (Fishtree, 2016).

Growth Engineering, este sistema ha estado innovando mediante la gamificación de su interfaz, y su uso educativo se basa en equipos, tableros de logros, retos y mucha interacción social (Growth Engineering, 2016).

Fidelis Education, el objetivo de esta plataforma es reducir la deserción universitaria acompañando al alumno mediante las relaciones humanas con sus mentores o tutores. Incorpora estrategias de gamificación y está vinculado con redes sociales corporativas como *LinkedIn*, de esta forma obtiene que los empleadores puedan hacer recomendaciones en que campos del desarrollo los estudiantes deben mejorar y así ellos puedan desarrollar las competencias que se requieren para el campo laboral (Fidelis, 2016).

Dreambox, este SGA combina una experiencia de aprendizaje altamente personalizada de las matemáticas con un plan de estudios riguroso para la comprensión profunda de los conceptos. Este software captura cada actividad realizada por los alumnos y puede anticipar 60 parámetros distintos de comportamiento (por ejemplo, frecuencia, tipo y velocidad de respuestas, cantidad y tipos de errores, entre otros). El programa amasa una cantidad inmensa de datos por alumno. Con esta información cambia la presentación, el tipo de clases y la secuencia siguiente en tiempo real ante cada alumno (DreamBox, 2016).

Smart Parrow, este software es de la Universidad de South Wales en Australia. Ofrece una interfaz adaptada para que escuelas y docentes creen clases adaptativas, simulaciones de ejercicios y tareas inteligentes. Este SGA está especialmente direccionado a docentes, debido a que les ofrece la autoría de las clases con un software que les permite

crear trayectos adaptativos. Tiene un enfoque centrado en aprender haciendo más que en la memorización (Smart Parrow, 2016).

2.4 Extracción de conocimiento en bases de datos

El descubrimiento de conocimiento implica diversos conceptos con los que está relacionado, tales como los Datos, la Información, la Gestión de la información y la gestión del conocimiento. Los datos consisten en hechos, imágenes, o sonidos. Cuando se combinan con la interpretación y el significado, se convierte en información. Por lo tanto, la información consiste en datos que han sido formateados, filtrados, y resumidos (Chen, 2001). El proceso de la transformación de datos en información se representa en el siguiente esquema (Itmazi, 2005).



Figura 4. Transformación de datos en información.

Una vez establecidos los conceptos datos e información, se puede definir del término de Gestión de la Información (GI), como “el proceso que utiliza herramientas y técnicas para capturar, recoger, analizar, organizar, manipular y recuperar la información. De la misma manera, controla y explota los recursos de la información de una organización” (Itmazi, 2005).

Por otro lado, la Gestión del Conocimiento (GC), se define como una disciplina emergente cuyo objetivo es generar, compartir y utilizar el conocimiento tácito y explícito

existente en un determinado espacio para dar respuesta a las necesidades de los individuos y de las comunidades en su desarrollo (Alavi et al., 2001).

Finalmente el Descubrimiento del Conocimiento en Bases de Datos (en inglés, *Knowledge Discovery in Databases*, KDD), se define como “el proceso no trivial de identificar patrones válidos, novedosos y potencialmente útiles y en última instancia, comprensible a partir de los datos” (Fayyad et al., 1996).

El KDD es un proceso iterativo e interactivo y de acuerdo con Hernández (Hernández et al., 2004) su taxonomía se puede organizar en cinco fases:

1. Integración y recopilación. En esta actividad se determinan las fuentes de información que pueden ser útiles y dónde conseguirlas. Posteriormente, se transforman los datos a un formato común. Dado que los datos provienen de diferentes fuentes, pueden contener valores erróneos o faltantes.
2. Selección, limpieza y transformación. Se eliminan o corrigen los datos incorrectos y se decide la estrategia a seguir con los datos incompletos. Además, se proyectan los datos para conseguir únicamente aquellas variables o atributos que van a ser relevantes, con el objetivo de hacer más fácil la tarea propia de minería y para que los resultados de la misma sean más útiles.
3. Minería de datos. Se decide cuál es la tarea a realizar (clasificar, agrupar, entre otras) y se elige el método a utilizar.
4. Evaluación e interpretación. Los expertos evalúan y se analizan los patrones, y si es necesario se vuelve a las fases anteriores para una nueva iteración. Esto

incluye resolver posibles conflictos con el conocimiento que se disponía anteriormente.

5. Difusión y uso. Se hace uso del nuevo conocimiento y se hace partícipe de él a todos los posibles usuarios. Este conocimiento se suele utilizar en procesos de toma de decisiones.

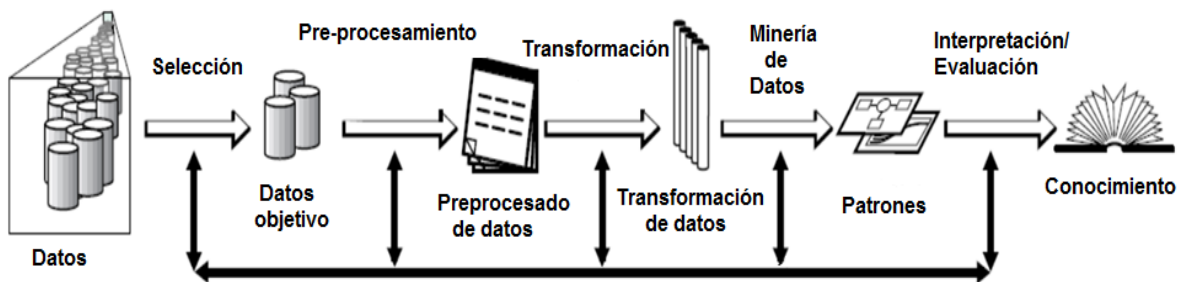


Figura 5. Proceso para la extracción de conocimiento en bases de datos (Fayyad, 1996).

Como se observa en la Figura 5, la minería de datos es una etapa crucial de las técnicas incluidas en el proceso KDD. Se define como “el proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos” (Fayyad et al., 1996).

La experimentación con las técnicas de minería de datos será una de las partes fundamentales para la extracción de conocimiento en esta investigación. Es por ello, que este documento se extenderá en este tema.

2.5 Minería de datos

La minería de datos, trabaja con grandes volúmenes de datos, procedentes en su mayoría de sistemas de información, con los problemas que ello conlleva (ruido, datos

ausentes, intratabilidad, volatilidad de los datos, entre otros), y aplica técnicas adecuadas para analizar estos datos y extraer conocimiento novedoso y útil (Hernández et al., 2004).

Otra definición importante en la literatura es la de Chen (2006), el cual señala que se asocia a bases de datos, en donde a través de diversas técnicas, se extrae información para generar conocimiento, el cual puede ser expresado a través de conceptos, reglas, leyes, patrones, entre otros.

La minería de datos se relacionada con diversas disciplinas emergentes, tales como: aprendizaje automático, aprendizaje profundo, reconocimiento de patrones, recuperación de información, aprendizaje estadístico, inteligencia artificial, entre otras.

2.5.1 Técnicas de minería de datos

Como ya se ha mencionado, la minería de datos consiste en el proceso en donde se extrae la información para descubrir patrones, y conocimiento, para ello requiere de diversas técnicas, de acuerdo con Romero (2006), estas son:

1. **Clasificación:** Consiste en determinar nuevos patrones, con base en un conjunto de patrones previamente etiquetados; método de aprendizaje supervisado
Algunos de los algoritmos más usados para ésta técnica son: ID3, J48, C4.5, Naive Bayes, Algoritmos Evolutivos, entre otros.
2. **Agrupamiento:** Su objetivo es agrupar datos de acuerdo a características similares, para ello, se analizan los datos almacenados en la base de datos, y de acuerdo a reglas de clasificación, se despliega una colección de objetos físicos o abstractos agrupados en clases. La agrupación se basa de la distancia de una dimensión n , donde n es el número de variables; se considera como un método

no supervisado. Algunos algoritmos representativos de esta técnica son: single-link, complete-link, k medias, EM (Expectation Maximization), entre otros.

3. Asociación: Su objetivo principal es establecer reglas que asocien conceptos de atributos diferentes de una misma base de datos. La asociación y correlación son usados para la búsqueda de un frecuente ítem de entre una gran cantidad de información. Lo anterior puede servir para el análisis del comportamiento de un estudiante. Algunos algoritmos representativos de esta técnica son: Apriori, Predictive A priori, entre otros.

2.5.2 Minería de datos en educación

La Minería de Datos Educativos (MDE, en inglés, *Educational Data Mining, EDM*), es una emergente y multidisciplinaria área de conocimiento, tiene como objetivo el desarrollo de métodos para la exploración de datos en un contexto educativo (Suhirman et al., 2014).

Existen diferencias significativas entre la minería de datos y la minería de datos educativos, en primer lugar, el primero es general, mientras que el segundo se especializa en datos provenientes de tecnologías educativas, por ejemplo los Sistemas de Gestión de Aprendizaje.

De acuerdo con Romero (2006), el objetivo del MDE depende del entorno, y a quien está dirigido. Para el primero importa si la enseñanza es presencial o distancia, mientras que para el segundo, si se tratan de alumnos, profesores o instituciones educativas (figura 6). Desde el punto de vista de los estudiantes, la minería de datos educativos puede determinar qué actividades, recursos y tareas, podrían mejorar su rendimiento, por otro

lado, desde el punto de vista del profesor, la MDE, tiene el potencial de evaluar y mejorar la didáctica del profesor; finalmente, desde el punto de vista de las instituciones educativas, permitirá optimizar las decisiones correspondientes al sistema, por ejemplo, mejorar la eficiencia del sitio web.

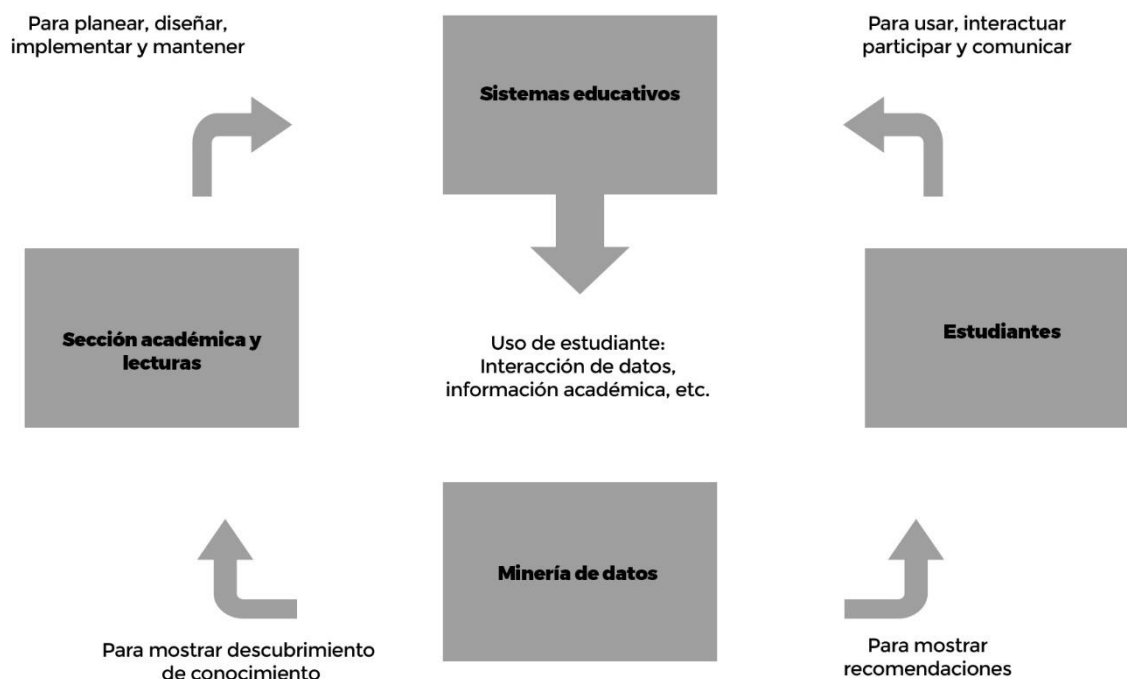


Figura 6. Participantes en la Minería de datos educativos.

En concreto, Suhirman et al. (2014) identifica a la MDE como aquella que orienta y mejora el proceso de aprendizaje de los estudiantes.

Como se ha señalado, la minería de datos provee beneficios a la educación, Wang (2014), manifiesta que es posible analizar la administración de la clase, a través de la información de los estudiantes, como por ejemplo, situación familiar, desempeño académico, número de ausencia del alumno en un mes, entre otros; asimismo se puede entrenar los talentos personales, analizar el impacto académico y la enseñanza.

Por otro lado Mavrikis (2013), señala que es posible predecir el desempeño académico del estudiante al estimar un valor desconocido de la variable que describe al estudiante, el cual puede ser numérico (regresión) o categórico (clasificación). El análisis de regresión permite relacionar las variables dependientes e independientes. La predicción es una de las aplicaciones más antiguas y populares en minería de datos aplicados a la educación.

De igual forma asegura que es posible predecir el riesgo de deserción al analizar el comportamiento del estudiante, además, que es posible una retroalimentación inteligente por parte de los SGA con el propósito de mejorar el desempeño académico, es posible recomendar cursos, estimar las destrezas del estudiante y detección del comportamiento a través de actividades basadas en comunidad y juegos.

Por otro lado, existe una variedad de paquetes software para implementar técnicas de minería de datos en los entornos educativos. Cada uno de ellos posee características apropiadas para realizar determinadas tareas o para analizar cierto tipo de datos. A continuación se describen algunos ejemplos:

- *WEKA (Waikato Environment for Knowledge Analysis)* (Witten, 2011), es una herramienta visual de libre distribución desarrollada por los investigadores de la Universidad de Waikato en Nueva Zelanda. Está implementado en el lenguaje de programación Java y opera en diferentes entornos de sistemas operativos. Algunas de sus principales características son: acceso a datos, preprocesado de datos, modelos de aprendizaje y contiene cuatro entornos.

- *SPSS (Statistical Product and Service Solutions)* Sistema que permite realizar análisis estadístico y gestión de información, es capaz de trabajar con datos procedentes de diversos formatos, como .XLSX (excel), genera desde gráficos y estadística descriptiva hasta estadísticos complejos. El sistema funciona en el sistema operativo Windows.
- *KEEL (Knowledge Extraction based on Evolutionary Learning)* (Alcalá-Fdez et al., 2008), fue desarrollado por 5 grupos de Universidades españolas, permite utilizar y construir diferentes modelos para minería de datos y tiene como característica importante la inclusión de una librería de algoritmos de aprendizaje evolutivo con código abierto en Java. Las características principales son: algoritmos de preprocesamiento, librería de algoritmos de extracción de conocimiento, librería de herramientas estadísticas para el análisis de los algoritmos y aplicación para su uso vía Web; además contiene cuatro entornos.
- *DBMiner* (DBMiner Technology Inc., 2003), es un software de distribución libre desarrollado por la Universidad de Simon Fraser, Canadá. Está concebido para la extracción del conocimiento en bases de datos relacionales, almacenes de datos y Web, su arquitectura de diseño incorpora OLAP (online analytic processing) y OLAM (online analytic mining).
- *RapidMiner* (Universidad de Dortmund, 2001). Este sistema antes era conocido como YALE (Yet Another Learning Environment), está desarrollado en Java y es de código abierto. Cuenta con una interfaz gráfica fácil de utilizar y además integra un mecanismo sencillo para desarrollar extensiones que hacen posible

integrar nuevos operadores y con ello adaptar el software para los requerimientos personales.

- *GISMO (Graphical Interactive Student Monitoring Tool for MOODLE)*, es una herramienta de monitorización gráfica interactiva que proporciona una visualización útil de las actividades de los estudiantes en los cursos en línea a los instructores. Con esta herramienta los instructores pueden examinar diversos aspectos de los estudiantes a distancia, tales como la asistencia a los cursos, la lectura de los materiales, la presentación de las asignaciones, entre otras (GISMO, 2016).

2.6 Investigaciones relacionadas

La MDE se ha aplicado con éxito tanto en entornos presenciales como virtuales de aprendizaje. A continuación se describen algunas investigaciones destacadas en esta área del conocimiento y que se encuentran relacionadas con la línea de investigación que se plantea.

La investigación de Romero, Ventura y García (2008), tuvo como propósito aplicar diversas técnicas de minería de datos en las actividades que realizaron alumnos de nivel superior dentro del entorno de un SGA basado en MOODLE. En cuanto a los métodos o técnicas de análisis de datos empleados, los autores recurrieron a la experimentación de técnicas de agrupamiento, reglas de asociación y se utilizaron diversos programas, tales como *WEKA*, *GISMO* y *KEEL*. La aportación de esta propuesta fue el desarrollo de un sistema que permitía a los profesores modificar las actividades del curso, y detección a tiempo si un estudiante está en posibilidad de fracaso.

Otra investigación conducida de igual manera por Romero, en conjunto con Espejo, Zafra, Romero y Ventura (2010), tuvo como propósito predecir la calificación final de estudiantes que emplean el Sistema de Gestión de Aprendizaje MOODLE. Para ello, utilizaron técnicas de clasificación.

La investigación de (Valsamidis et al., 2012), tuvo como objetivo experimentar diversas técnicas de agrupamiento para analizar las actividades de los estudiantes de pregrado dentro del entorno de un SGA. Para ello, se utilizaron los algoritmos de cadenas de *Markov* y *SimpleKMeans* a través del software *WEKA*. Adicionalmente se incluyó una visualización de los resultados a través del uso del software *3D graphs*.

Hung, Hsu y Rice (2012), desarrollaron un estudio cuyo propósito fue la evaluación de un programa educativo K-12. Para el estudio se aplicaron algoritmos de agrupamiento. En cuanto al software utilizaron *SAS Enterprise Miner 6.1*. Los principales resultados de la investigación fueron la caracterización completa de los alumnos de educación básica.

Una investigación de importancia, fue la realizada por (Gobert et al., 2013). El objetivo del estudio fue el desarrollo de un método para evaluar las habilidades científicas de estudiantes de nivel superior. Para ello, implementaron técnicas de clasificación de minería de datos y utilizaron el software *RapidMiner* versión 4.6.

Un trabajo destacado es el (You, 2015), el cual condujo una investigación cuyo objetivo fue examinar el efecto de la procrastinación de estudiantes, usando datos de un SGA a través de la implementación de minería de datos. El autor recurrió a la estadística descriptiva y el análisis de correlación. El principal resultado fue: El ausentismo escolar y

el retraso en la entrega de tareas, son los principales factores que originan la procrastinación y se correlacionan negativamente con la calificación final del curso.

El autor (Magdin, 2015), desarrollo una investigación cuyo objetivo fue personalizar recomendaciones de cursos en línea a estudiantes. Para el estudio se aplicaron reglas de asociación. Los softwares empleados fueron *Module Interactive Element Stat* y *Microsoft Excel*. Los principales resultados de esta investigación fueron la implementó un cuestionario que, de acuerdo a reglas de asociación, permitió la recomendación de materiales específicos para las preferencias de cada estudiante y a pesar de la recomendación de material, los estudiantes continuaron mostrando deficiencias en el análisis de las lecturas. El estudio concluyó que los estudiantes activos están más cómodos con actividades basadas en problemas y discusión en grupo, mientras que los estudiantes reflexivos prefieren más las lecturas, por consiguiente no es fácil determinar una estrategia de enseñanza estándar.

Los autores (Yildiz & Gulsecen, 2015), desarrollaron una investigación basada en el desempeño del estudiante para predecir la probabilidad de su abandono escolar. Para ello, recurrieron a técnicas de agrupamiento y utilizaron el software Matlab.

Finalmente, se describe el trabajo de (Nor et al., 2015), quienes desarrollaron una investigación cuyo objetivo se centraba en entender el comportamiento del estudiante en relación con la búsqueda de información y materiales en un sistema educativo basado en la web, para la generación de una recomendación personalizada de contenido. Los autores recurrieron a técnicas de agrupamiento y utilizaron el software *WEKA*.

De acuerdo con la revisión de nueve trabajos relacionados con el tema de investigación que se está planteando, se detectó que la implementación de técnicas de agrupamiento son las más recurrentes. Cabe destacar que tres de las nueve investigaciones utilizaron alguna técnica de minería de datos para realizar predicciones del desempeño académico de los estudiantes. Asimismo, tres de las nueve investigaciones utilizaron el software *WEKA*. Otro elemento a destacar es que dos de las investigaciones descritas desarrollaron su propio software.

A continuación se presenta un cuadro de análisis comparativo que muestra los aspectos comunes entre las investigaciones presentadas anteriormente. Para la comprensión de dicho cuadro (tabla 1), se deben considerar los siguientes términos empleados:

1. Finalidad: esta categoría identifica la finalidad de la investigación.
2. Sistemas de Gestión del Aprendizaje: se refiere al sistema del cual se extrajo la información.
3. Software: representa el software empleado para el análisis de datos.

Tabla 1.

Comparación de las investigaciones relacionadas.

Trabajos relacionados	Finalidad	SGA	Software
Romero, et.al (2008)	Predecir	MOODLE	WEKA
Romero, et.al (2010)	Predecir	Otras	Software Propio
Valsamidis, et. al (2012)	Recomendar	Otras	WEKA y Otros

Tabla 1 (continuación)

Trabajos relacionados	Finalidad	SGA	Software
Hung, Hsu y Rice (2012)	Evaluar	Otras	Software Propio
Gobert, et. al (2013)	Desarrollar	Otras	Otros
You (2015)	Evaluar	Otras	Otros
Magdin (2015)	Recomendar	MOODLE	Otros
Yildiz y Gulsecen (2015)	Predecir y Evaluar	MOODLE	Otros
Nor, et. al (2015)	Evaluar y Recomendar	Otras	WEKA

En la tabla 1 se expone una comparación de trabajos relacionados con la presente investigación. En ella se puede apreciar una comparación de aspectos que interesan a éste trabajo, tales como la finalidad del estudio, el software y el SGA utilizado.

Capítulo 3. Metodología

El presente capítulo tiene como propósito describir el procedimiento efectuado para el cumplimiento de la investigación.

Respecto al alcance de la investigación, se declara que la tesis se sustenta bajo un paradigma positivista, el investigador reconoce una realidad independiente a éste y capaz de ser medible; es de enfoque cuantitativo debido al uso de datos cuantitativos y técnicas de análisis cuantitativas; y exploratoria, ya que es primera de su tipo (Hernández, 2006).

De acuerdo al autor Kerlinger (1969), este estudio es considerado como no experimental debido a que no se manipularon las variables, ni se asignaron aleatoriamente a los sujetos estudiados, los cuales fueron todos los profesores que impartieron al menos un curso en UADY Virtual; asimismo, es *ex post-facto* ya que los eventos, al momento de la investigación, ya habían ocurrido al igual que sus efectos.

En cuanto a la temporalidad, de acuerdo con León y Montero (1997), el proyecto fue transversal, ya que se desarrolló en un momento determinado y debido a que permitió establecer la relación entre las variables de interés.

La metodología empleada en la investigación fue descubrimiento de conocimiento en bases de datos (en inglés, *Knowledge Discovery from Dataset*, KDD), la cual consiste en cinco fases: 1. Selección, 2. Preproceso, 3. Transformación, 4. Minería de datos y 5. Análisis (ver Figura 7). A continuación se describen las acciones ejecutadas en cada una de las fases.

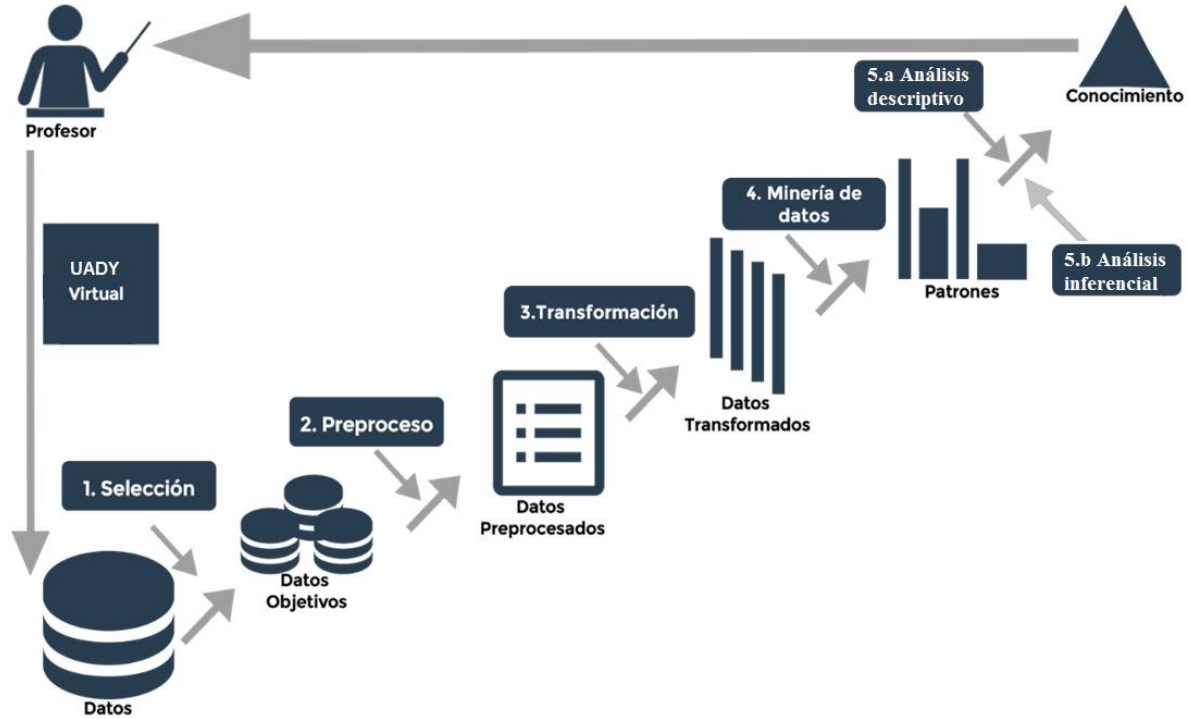


Figura 7. Metodología KDD.

3.1 Fase 1: Selección

El Departamento de Innovación e Investigación Educativa (DIIE) de la Dirección General de Desarrollo Académico (DGDA) de la UADY, proporcionó las bases de datos que contenían toda la actividad de los profesores realizada en el entorno de UADY Virtual (<http://uadyvirtualcloud.uady.mx/>) durante el período enero – julio del año 2016.

Adicionalmente, facilitó información complementaria relacionada con la habilitación de los profesores en el uso de las herramientas que posee este sistema.

Los registros de acceso de los profesores estaban distribuidos en 197 archivos (144 MB) en formato CSV (del inglés, *Comma Separated Value*) (ver Figura 8).

Archivos actualmente en el disco (197)

logs_20160715-0846-20160715_1346-comma_separated	15/07/2016 09:07 a...	Archivo de valores...	90 KB
logs_20160715-0847-20160715_1347-comma_separated	15/07/2016 09:08 a...	Archivo de valores...	1,298 KB
logs_20160715-0848-20160715_1348-comma_separated	15/07/2016 09:08 a...	Archivo de valores...	1,287 KB
logs_20160715-0849-20160715_1349-comma_separated (1)	15/07/2016 09:09 a...	Archivo de valores...	2,001 KB
logs_20160715-0849-20160715_1349-comma_separated	15/07/2016 09:10 a...	Archivo de valores...	1,540 KB
logs_20160715-0850-20160715_1350-comma_separated	15/07/2016 09:14 a...	Archivo de valores...	324 KB
logs_20160715-0855-20160715_1355-comma_separated	15/07/2016 09:15 a...	Archivo de valores...	116 KB
logs_20160715-0856-20160715_1356-comma_separated	15/07/2016 09:15 a...	Archivo de valores...	1,165 KB
logs_20160715-0857-20160715_1357-comma_separated	15/07/2016 09:16 a...	Archivo de valores...	1,482 KB
logs_20160715-0859-20160715_1359-comma_separated	15/07/2016 09:16 a...	Archivo de valores...	1,461 KB
logs_20160715-0900-20160715_1400-comma_separated	15/07/2016 09:17 a...	Archivo de valores...	1,181 KB
logs_20160715-0901-20160715_1401-comma_separated (1)	15/07/2016 09:17 a...	Archivo de valores...	507 KB
logs_20160715-0901-20160715_1401-comma_separated	15/07/2016 09:18 a...	Archivo de valores...	1,310 KB
logs_20160715-0902-20160715_1402-comma_separated	15/07/2016 09:18 a...	Archivo de valores...	78 KB
logs_20160715-0903-20160715_1403-comma_separated	15/07/2016 09:18 a...	Archivo de valores...	871 KB
logs_20160715-0919-20160715_1419-comma_separated	15/07/2016 10:19 a...	Archivo de valores...	479 KB
logs_20160715-0920-20160715_1420-comma_separated (1)	15/07/2016 10:20 a...	Archivo de valores...	1,151 KB
logs_20160715-0920-20160715_1420-comma_separated	15/07/2016 10:20 a...	Archivo de valores...	718 KB
logs_20160715-0921-20160715_1421-comma_separated (1)	15/07/2016 10:21 a...	Archivo de valores...	282 KB
logs_20160715-0921-20160715_1421-comma_separated	15/07/2016 10:21 a...	Archivo de valores...	1,685 KB
logs_20160715-0923-20160715_1423-comma_separated (1)	15/07/2016 10:22 a...	Archivo de valores...	532 KB
logs_20160715-0923-20160715_1423-comma_separated	15/07/2016 10:22 a...	Archivo de valores...	156 KB

Figura 8. Imagen parcial de las bases de datos que contenían los registros de acceso de los profesores en UADY Virtual.

Cada base de datos estaba compuesta por siete atributos los cuales se describen en la siguiente tabla:

Tabla 2.

Atributos de las bases de datos.

#	Atributo	Descripción
1	Hora	Registra el momento exacto en el que se crea el archivo de registro.
2	Contexto del evento	Almacena la dirección web que describe la ruta del elemento seleccionado.
3	Componente	Identifica si la dirección web corresponde a una actividad o recurso del sistema.
4	Nombre del evento	Muestra la dirección web de la acción del elemento seleccionado.

Tabla 2 (continuación)

#	Atributo	Descripción
5	Descripción	Identifica al usuario que efectuó alguna acción y describe lo que realizó.
6	Origen	Señala el origen del archivo de registro.
7	Dirección IP (<i>Internet Protocol</i>)	Identifica la dirección IP del usuario que accedió al sistema.

Posteriormente se agruparon los 197 archivos .CSV en un archivo .XLSX, lo cual permitió consolidar toda la información en un mismo documento, con el propósito de obtener todos los registros de acceso de los profesores, y particularmente, determinar las herramientas del SGA usadas por éstos.

3.2 Fase 2: Preproceso

Esta es la fase que requirió la mayor cantidad de tiempo en el proceso de extracción del conocimiento. Su principal objetivo consistió en ubicar a los profesores a través de un Número de Identificación Único (en inglés, *Identification*, ID) y determinar las herramientas (así como la frecuencia de uso de éstas) empleadas para cada uno de los profesores. Para ello, se utilizó el archivo .XLSX creado en la fase anterior, y el software *Excel* versión 2010.

Antes de describir las acciones realizadas en el preproceso, conviene en primera instancia definir las herramientas que pueden emplear los profesores en el entorno de UADY Virtual:

1. Asistencia: mantiene un registro en línea de la asistencia de los estudiantes.
2. Base de datos: permite tanto al profesor como al estudiante, mostrar y buscar un banco de entradas de registros acerca de cualquier tópico concebible. El formato

y la estructura de estas entradas pueden ser casi ilimitadas, incluyendo a imágenes, archivos, URL, números y texto, entre otras.

3. Chat: permite a los participantes pueden tener una discusión sincrónica en tiempo real.
4. Consulta: sirve para que el profesor formule una pregunta y para ello, especifica una serie de respuestas entre las cuales deben elegir los alumnos.
5. Cuestionario: permite encuestar a los participantes de un curso.
6. Encuesta: permite realizar encuestas.
7. Encuesta predefinida: proporciona varios instrumentos verificados, incluyendo COLLES (*Constructivist On-Line Learning Environment Survey*) y ATTLS (*Attitudes to Thinking and Learning Survey*), que han sido encontrado útiles en la evaluación y estimulación del aprendizaje en ambientes en línea. Los maestros pueden usar estas actividades para recolectar datos de sus alumnos, que les ayudarán a conocerlos y reflexionar sobre su propia enseñanza.
8. Foro: los estudiantes y profesores pueden intercambiar ideas al publicar comentarios como parte de un 'hilo' de una discusión. Se pueden incluir archivos tales como imágenes. El profesor puede elegir valorar publicaciones en foros y también es posible otorgar permiso a los estudiantes para que valoren las publicaciones de unos a otros.
9. Glosario: los participantes pueden crear y mantener una lista de definiciones, similar a un diccionario.

10. *Hot Potatoes*: permite al profesor elaborar una diversidad de ejercicios, tales como: crucigramas, rompecabezas, entre otros. Una vez finalizados se incorporan al curso.
11. Lecciones: presenta una serie de páginas en formato HTML (*HyperText Markup Language*) al estudiante, a quien generalmente se le pregunta que elija una opción entre varias al final del contenido. La opción que elija le llevará a una página específica de la lección.
12. SCORM: permite subir cualquier paquete SCORM o AICC para incluirlo en su curso.
13. Taller: es una actividad de evaluación entre pares con varias opciones.
14. Tareas: permite a los profesores recolectar trabajos de sus estudiantes, revisarlos y proporcionarles retroalimentación, incluyendo calificaciones. El trabajo que envían los alumnos es visible solamente para el profesor a menos que se hubiera elegido una tarea de grupo.
15. Wiki: es una colección de documentos web escritos en forma colaborativa.
16. Archivo: proporciona una forma fácil para que un profesor exponga materiales a sus estudiantes. Por ejemplo, documentos de procesador de texto o presentaciones de diapositivas. Pueden subirse todos los tipos de archivos, pero los estudiantes necesitan tener el programa correcto para poder abrirlos.
17. Carpeta: permite al profesor mostrar varios recursos de un curso juntos. Los recursos pueden ser de tipos diferentes y pueden subirse agrupados, por ejemplo,

en un archivo comprimido ZIP, el cual es expandido posteriormente, o pueden añadirse de uno a uno hacia una carpeta vacía en la página del curso.

18. Etiqueta: sirve como un espaciador dentro de una página. Puede usarse para añadir texto, imágenes, multimedia o código, entre otros recursos en diferentes secciones. Es un recurso muy versátil y puede ayudar a mejorar la apariencia de un curso si se usa inteligentemente. Se pueden añadir banners o descripciones para diferenciar y resaltar áreas diferentes.
19. Libro: permite crear recursos de diversas páginas con un formato similar al de un libro. La información se puede organizar en capítulos principales y sub capítulos.
20. Página: crea un enlace hacia una pantalla que muestra el contenido creado por el profesor. El robusto editor de texto permite que la página muestre muchos tipos diferentes de contenidos, tales como texto plano, imágenes, audio, video, código incrustado o una combinación de éstos.
21. URL (*Uniform or Universal Resource Locator*): es un enlace hacia un sitio web. Los profesores pueden usar este recurso para proporcionar a sus estudiantes enlaces web para investigación, ahorrándoles tiempo y esfuerzo. Los URL pueden mostrarse en varias maneras, por ejemplo, abrir en una nueva ventana, de forma que el estudiante pueda acceder y usar el URL y después cerrarlo y regresar con facilidad a su página original del curso Moodle.

Con el objetivo de reconocer las herramientas que utilizó cada profesor en el entorno de UADY Virtual, se seleccionaron los atributos Nombre del evento y Descripción.

Posteriormente, se realizó la limpieza de datos para identificar la frecuencia de uso para cada herramienta utilizada por el profesor. Para ello, se realizaron las siguientes acciones:

1. Se filtraron los archivos de registro considerando únicamente las herramientas que pueden usar los profesores, las cuales se obtuvieron del atributo “Nombre del evento”.
2. Se creó la primera versión de la base de datos que relacionaba el ID del profesor con la herramienta utilizada. La identificación se obtuvo con el atributo “Descripción”,
3. Se contabilizaron cada una de las herramientas con la función Subtotal del programa *Excel* (ver figura 9).

	A	B	C
1	User id	Activity	
2		2 workshop	
3		2 workshop	
6		20 assign	
7		20 assign	
8		20 assign	
9		20 assign	
10		20 assign	

Figura 9. Vista parcial del uso de la función Subtotal obtenida con el software *Excel*.

4. Se creó la segunda versión de la base de datos en la que se relacionaba el ID del profesor y los archivos de acceso correspondientes para cada herramienta usada por cada profesor (ver figura 10).

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	id	assignments	chat	choice	database	forum	glossary	lesson	quiz	SCORM	survey	wiki	workshop
2	20	1253	0	40	2	137	16	9	900	10	8	1	70
3	114	5664	4	44	22	134	13	15	796	1	3	2	153
4	428	7289	3	42	37	99	24	10	536	39	4	9	143
5	1400	333	6	43	17	110	23	7	722	20	1	5	163
6	1799	5154	6	0	6	51	2	10	845	21	2	1	20
7	1800	4219	5	44	26	119	1	1	517	38	6	6	17
8	1802	2737	2	24	7	27	26	7	78	2	1	4	174
9	1805	1510	0	40	31	39	26	20	569	18	2	9	123
10	1808	7254	3	0	13	53	22	20	895	11	12	5	83
11	1810	7203	2	46	17	19	26	2	389	45	3	4	47
12	1811	5468	4	45	20	66	16	13	105	31	2	9	112
13	1812	2016	3	18	39	56	6	16	758	21	1	7	71
14	1814	6428	3	21	0	53	6	5	133	28	11	8	174
15	1815	3721	3	43	10	66	16	10	262	26	8	4	43
16	1816	3092	4	37	39	103	12	11	200	23	12	1	10
17	1819	3023	2	28	24	126	11	14	804	27	6	4	179
18	1824	1355	4	18	34	122	10	8	133	35	5	4	132
19	1825	398	2	23	33	79	3	19	227	11	6	4	89
20	1853	4240	5	28	38	52	23	20	539	33	11	9	27
21	1870	4080	3	15	38	84	7	19	396	27	1	2	116
22	1923	4436	3	4	25	52	1	20	372	13	12	8	33
23	1925	4440	4	25	13	68	16	10	258	36	4	3	132
24	1927	3006	2	44	13	7	1	20	170	25	7	9	4
25	1928	446	1	15	32	53	25	11	387	45	9	8	88

Figura 10. Vista parcial de la base de datos UADY Virtual V2.0.

- Se complementó la información de la base de datos UADY Virtual V2.0 con tres datos adicionales, el primero y el más importante, la Facultad a la que correspondía el número de identificación del profesor y si éste aprobó el módulo de capacitación en el uso de UADY Virtual impartido por el personal del MEFI (ver figura 11).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Id	Facultad	Curso	Acreditado	Acreditado	assignments	chat	choice	database	forum	glossary	lesson	quiz	SCORM	survey	wiki	workshop
2	20	Ingeniería	No	No	No	1253	0	40	2	137	16	9	900	10	8	1	70
3	114	Ingeniería	No	No	No	5664	4	44	22	134	13	15	796	1	3	2	153
4	428	Contaduría y	Si	Si	dic-15	7289	3	42	37	99	24	10	536	39	4	9	143
5	1400	Contaduría y	Si	Si	dic-14	333	6	43	17	110	23	7	722	20	1	5	163
6	1799	Medicina Vet	Si	Si	jul-15	5154	6	0	6	51	2	10	845	21	2	1	20
7	1800	Ingeniería Qu	Si	Si	dic-14	4219	5	44	26	119	1	1	517	38	6	6	17
8	1802	Matemáticas	Si	Si	No	2737	2	24	7	27	26	7	78	2	1	4	174
9	1805	Ingeniería	Si	Si	nov-14	1510	0	40	31	39	26	20	569	18	2	9	123
10	1808	Enfermería	Si	Si	No	7254	3	0	13	53	22	20	895	11	12	5	83
11	1810	Contaduría y	Si	Si	oct-14	7203	2	46	17	19	26	2	389	45	3	4	47
12	1811	Contaduría y	Si	Si	jun-14	5468	4	45	20	66	16	13	105	31	2	9	112
13	1812	Educación	Si	Si	mar-13	2016	3	18	39	56	6	16	758	21	1	7	71
14	1814	Educación	Si	Si	mar-13	6428	3	21	0	53	6	5	133	28	11	8	174
15	1815	Educación	Si	Si	jun-14	3721	3	43	10	66	16	10	262	26	8	4	43
16	1816	Derecho	Si	Si	No	3092	4	37	39	103	12	11	200	23	12	1	10
17	1819	Arquitectura	Si	Si	No	3023	2	28	24	126	11	14	804	27	6	4	179
18	1824	Matemáticas	Si	Si	No	1355	4	18	34	122	10	8	133	35	5	4	132
19	1825	Química	Si	Si	No	398	2	23	33	79	3	19	227	11	6	4	89
20	1853	Enfermería	No	No	No	4240	5	28	38	52	23	20	539	33	11	9	27
21	1870	Química	Si	Si	jun-13	4080	3	15	38	84	7	19	396	27	1	2	116
22	1923	Derecho	Si	Si	nov-14	4436	3	4	25	52	1	20	372	13	12	8	33
23	1925	Derecho	Si	Si	nov-14	4440	4	25	13	68	16	10	258	36	4	3	132
24	1927	Derecho	Si	Si	jun-14	3006	2	44	13	7	1	20	170	25	7	9	4
25	1928	Derecho	Si	Si	jun-14	446	1	15	32	53	25	11	387	45	9	8	88

Figura 11. Vista parcial de la base de datos UADY Virtual V2.0.

Posteriormente se modificaron los datos nominales del atributo Facultad, por Campus, para ello se agruparon a las facultades de acuerdo al área de conocimiento al cual pertenecen. Dichas áreas de conocimiento son:

1. Arquitectura, Hábitat, Arte y Diseño (AHAD).
2. Ciencias Biológicas y Agropecuarias (CBA).
3. Ciencias Exactas e Ingenierías (CEI).
4. Ciencias de la Salud (CS).
5. Ciencias Sociales Económico-Administrativas y Humanidades (CSEAH).

Con la acción anterior se dio finalizada la fase 2, dando como resultado la versión final de la base de datos (UADY Virtual V3.0).

3.3 Fase 3: Transformación

Adicional a la base de datos UADY Virtual V3.0, cuyos atributos eran numéricos, se creó una cuarta versión de la base de datos en donde se transformó los valores de las herramientas de numérico a nominales; los posibles valores nominales fueron: Muy bajo, Bajo, Medio, Alto y Muy alto, los cuales representan la calificación de la frecuencia de uso, para cada herramienta y para cada profesor.

Para determinar los nuevos valores nominales, se identificaron los valores máximos y mínimos de cada herramienta; posteriormente, se calculó el rango (valor máximo menos el valor mínimo, dividido entre cinco) para conformar la escala Likert. Como resultado de lo anterior se finalizó la creación de la base de datos UADY Virtual V4.0 (ver figura 12).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	ProfesorID	Campus	CursoMS	AcreditoMS	Fecha	Asistencia	BaseDatos	Chat	Consulta	Cuestionario	Encuesta	EncuestaPredefinida	Foro	Glosario	HotPotatoes	Lecciones	SCORM
20	CEI	No	No	No	Muy poco	Muy poco	Muy poco	Regular	Muy poco	Muy poco	Muy poco	Muy poco	Elevado	Muy poco	Muy poco	Muy poco	Normal
114	CEI	No	No	No	Muy poco	Muy poco	Muy poco	Elevado	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Elevado
428	CSEAH	Si	Si	12/2015	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco
1400	CSEAH	Si	Si	12/2014	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco
1799	CBA	Si	Si	07/2015	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco
1800	CEI	Si	Si	12/2014	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco
1802	CEI	Si	Si	ATI	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco
1805	CEI	Si	Si	11/2014	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco
1808	CS	Si	Si	ATI	Muy poco	Muy poco	Elevado	Muy poco	Regular	Muy poco	Muy poco	Muy poco	Muy poco	Regular	Muy poco	Muy poco	Muy poco
1810	CSEAH	Si	Si	10/2014	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco
1811	CSEAH	Si	Si	06/2014	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco
1812	CSEAH	Si	Si	03/2013	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco
1814	CSEAH	Si	Si	03/2013	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco
1815	CSEAH	Si	Si	06/2014	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Regular	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco
1816	CSEAH	Si	Si	ATI	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco
1819	AHAD	Si	Si	ATI	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco
1824	CEI	Si	Si	ATI	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco
1853	CS	No	No	No	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco
1870	CS	Si	Si	06/2013	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco
1923	CSEAH	Si	Si	11/2014	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco
1925	CSEAH	Si	Si	11/2014	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco
1927	CSEAH	Si	Si	06/2014	Muy poco	Muy poco	Frecuente	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco
1928	CSEAH	Si	Si	06/2014	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco
1990	CSEAH	Si	Si	ATI	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco
1992	CSEAH	Si	Si	06/2014	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco
1993	CSEAH	Si	Si	06/2014	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco
1995	CEI	Si	Si	11/2014	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco
1996	CEI	Si	Si	11/2014	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco	Muy poco

Figura 12. Base de datos UADY Virtual V4.0

Cabe destacar que se creó una fórmula con la función lógica SI, para automatizar el proceso de etiquetado de los valores. A continuación se muestra un ejemplo de la fórmula aplicada para el etiquetado del atributo Tareas.

$$\begin{aligned}
&=SI(Y(Etiquetas!\$B\$8<=Original!F2,Original!F2<Etiquetas!\$C\$8),Etiquetas!\$A\$8, \\
&SI(Y(Etiquetas!\$B\$9<=Original!F2,Original!F2<Etiquetas!\$C\$9),Etiquetas!\$A\$9, \\
&SI(Y(Etiquetas!\$B\$10<=Original!F2,Original!F2<Etiquetas!\$C\$10),Etiquetas!\$A\$10, \\
&SI(Y(Etiquetas!\$B\$11<=Original!F2,Original!F2<Etiquetas!\$C\$11),Etiquetas!\$A\$11, \\
&SI(Y(Etiquetas!\$B\$12<=Original!F2,Original!F2<=Etiquetas!\$C\$12),Etiquetas!\$A\$12,"No"))))
\end{aligned} \tag{1}$$

Finalmente se convirtió el tipo de archivo de las dos bases de datos (UADY Virtual V3.0 y V4.0) de .XLSX a .ARFF (del inglés, *Attribute-Relation File Format*).

3.4 Fase 4: Minería de datos

En esta esta etapa se aplicaron diversas técnicas de minería de datos a través de los programas WEKA y SPSS. A continuación se describen sus principales características:

WEKA

El software WEKA (*Waikato Environment for Knowledge Analysis*) consiste en un conjunto de algoritmos de aprendizaje automatizado, los cuales están basados en el lenguaje de programación JAVA; su propósito es aplicar tareas de minería de datos y extraer conocimiento de las bases de datos. El sistema posee una licencia de uso GNU (*General Public License*), lo cual lo convierte en un sistema de código abierto, lo que significa que los usuarios poseen la libertad de modificar el código fuente del mismo.

Una de las principales características del software es su interfaz visual, lo que facilita el uso del sistema. Es posible “cargar” bases de datos en diversas extensiones, de las cuales destaca la extensión .ARFF, motivo por el cual se transformaron las bases de datos en la fase anterior. El sistema se compone de cuatro entornos de trabajo, de acuerdo con Hernández y Ramírez (2006) se definen como:

1. *Explorer*: entorno visual para emplear los paquetes (algoritmos) del sistema.
2. *Experimenter*: diseñado para experimentos de gran escala, permite automatizar diversas tareas.
3. *Knowledge Flow*: basado en flujo de información, permite análisis de minería de datos.
4. *Simple CLI*: entorno consola que permite emplear los paquetes del sistema directamente con código.

Las versiones posteriores WEKA versión 3.4, ofrecen adicionalmente el entorno de trabajo *workbench*.

De los cuatro entornos de trabajo mencionados anteriormente, para el desarrollo de la investigación se utilizó el entorno de trabajo *visual explorer*. A continuación se describen sus principales características (Iglesias, Martínez, Aler, & Fernández, 2009):

- La versión 3.8 (versión que se utilizó para la investigación) se compone de seis pestañas: *Preprocess*, *Classify*, *Cluster*, *Associate*, *Select attributes* y *Visualize*.
- En el *Preprocess* es posible aplicar diversos filtros tanto de algoritmos supervisados como no supervisados.
- En el *Classify* es posible aplicar algoritmos supervisados (empleado para la creación de modelos predictivos), mientras que en el *Cluster* los no supervisados (empleado para la creación de los *clusters*).
- En el *Associate* es posible aplicar algoritmos de asociación.

SPSS

De acuerdo con Gómez y Menéndez (2011), este software es una herramienta informática cuyas funciones básicas se resumen en tres: 1. Leer datos de n instancias (renglones) en m atributos (columnas), 2. Aplicación de análisis estadístico y 3. Visualización de los resultados.

La estructura del programa se conforma principalmente por el editor de datos y el visor de resultados. El primero posee dos ventanas, ventana de datos, donde se almacenan los datos que serán procesados y la ventana de variables, donde se nombran y definen las variables. El segundo, el visor de resultados, muestra a través de estadísticos (texto), tablas y gráficos, los resultados de los análisis estadísticos derivados del editor de datos.

Con el SPSS se pueden realizar múltiples acciones y análisis estadísticos; es posible utilizar operadores aritméticos, relacionales y lógicos, aplicar funciones exponenciales, logarítmicas, trigonométricas, numéricas, de estadística descriptiva (moda, media, mediana), lógicas, de conversión, etc.; se pueden aplicar técnicas de regresión múltiple, análisis de discriminante, de agrupamiento (análisis de *cluster*), análisis factorial, por mencionar algunos (Pérez, 2005).

3.4.1 Agrupamiento

Para el agrupamiento se utilizó la base de datos UADY Virtual V3.0; el objetivo fue agrupar a los profesores respecto a su desempeño en el uso del sistema UADY Virtual de acuerdo a los valores: Muy alto, Alto, Medio, Bajo y Muy Bajo; Para ello, se consideró utilizar tanto el software SPSS como el WEKA con el objetivo de complementar los resultados.

3.4.2 Clasificación

Para el desarrollo de la técnica de clasificación se usó la base de datos derivada del análisis de agrupamiento (UADY Virtual V3.1), se consideró el atributo Grupo como la variable de respuesta (datos tipo nominales), y las herramientas más los atributos nominales CursoM5 y HabilitoM5 como las variables predictivas (datos tipo numéricos).

La clasificación se desarrolló con el software WEKA. Se utilizó el algoritmo J48 por su popularidad determinado de las investigaciones relacionadas (ver capítulo 2, estado del arte). El resultado de la aplicación de la técnica de clasificación se expone en el siguiente capítulo a manera de diagrama de árbol de decisión.

3.4.3 Asociación

Para la aplicación de la técnica de asociación se usó la base de datos UADY Virtual V4.0, únicamente se consideraron los atributos herramientas, ya que por método, en la asociación no se define una variable de respuesta.

Nuevamente se empleó el sistema WEKA. Se desarrolló la técnica con apoyo del algoritmo A priori, y se instruyó al ordenador mostrar las 10 reglas de asociación más relevantes. Los resultados se muestran en el capítulo posterior.

3.5 Fase 5: Análisis de los resultados de las técnicas de minería de datos

En esta última etapa de la metodología KDD, se analizaron los resultados generados de las tres técnicas de minería de datos descritas anteriormente (agrupamiento, clasificación y asociación), lo cual concluyó en la generación de nuevo conocimiento respecto al uso de UADY Virtual por parte de los profesores de esta institución. Los resultados y las conclusiones se presentan en los siguientes capítulos del presente documento.

3.6 Fase 6: Análisis descriptivo e inferencial.

Con el objeto de complementar el análisis de minería de datos, se desarrolló un análisis descriptivo a través de la creación de diversos indicadores, que permitieron identificar características generales del uso del sistema UADY Virtual por parte de los profesores; los indicadores se generaron a petición del DIIE.

Asimismo se aplicó un análisis inferencial para pruebas no paramétricas, a fin de comparar, y en su caso establecer si existieron diferencias significativas en el uso de UADY Virtual de acuerdo a las áreas de conocimiento. Se creó un capítulo exclusivo para la presentación de los resultados del análisis descriptivo e inferencial.

Capítulo 4. Resultados de la Minería de Datos

Como ya se ha mencionado en la metodología, se emplearon tres técnicas de minería de datos, agrupamiento, clasificación y asociación. El objetivo del agrupamiento fue crear grupos de profesores con características similares, por otro lado en la clasificación se desarrolló un modelo predictivo del desempeño docente, y finalmente, con la asociación se obtuvieron las correlaciones entre las herramientas de UADY Virtual. A continuación se presentan los resultados para cada técnica.

4.1 Técnicas de agrupamiento

Los resultados del análisis de agrupamiento a través del software SPSS versión 22 indicaron la existencia de cinco grupos (tabla 3), el primero con 1 profesor, el segundo con 409, el tercero con 52, el cuarto con 18 y el quinto con 2. Con el objetivo de identificar qué grupo correspondía a la calificación “Muy alto”, “Alto”, “Medio”, “Bajo” y “Muy bajo”, se procedió a registrar el porcentaje de las frecuencias para las variables nominales (CursoM5 y AcreditoM5) y las medias para las variables numéricas (todas las herramientas).

Tabla 3.

Características de los grupos de profesores.

Variable	Grupo 1 (1)	Grupo 2 (409)	Grupo 3 (52)	Grupo 4 (18)	Grupo 5 (2)
CursoM5	100%	86.6%	84.6%	88.9%	100%
AcreditoM5	100%	76.3%	82.7%	88.9%	100%
Asistencia	0.00	.03	.21	.28	0.00
Base de Datos	0.00	.18	.23	0.00	0.00
Chat	0.00	.07	.19	0.00	0.00
Consulta	0.00	10.36	28.54	89.72	0.00
Cuestionario	0.00	.51	.54	.44	0.00
Encuesta	0.00	.17	.27	.11	0.00
Encuesta Predefinida	0.00	.22	.15	0.00	0.00
Foro	0.00	2.84	5.60	13.22	0.00

Tabla 3 (continuación)

Variable	Grupo 1 (1)	Grupo 2 (409)	Grupo 3 (52)	Grupo 4 (18)	Grupo 5 (2)
Glosario	0.00	.40	.60	0.00	0.00
Hot Potatoes	0.00	.01	0.00	.11	0.00
Lecciones	0.00	.01	0.00	1.11	0.00
SCORM	0.00	.03	0.00	3.50	0.00
Taller	0.00	.10	.04	9.94	0.00
Tareas	5181.00	69.63	852.52	2182.78	7289.50
Wiki	0.00	.10	.17	0.00	0.00
Archivo	0.00	14.56	21.69	29.61	0.00
Carpeta	0.00	3.42	5.19	7.61	.50
Etiqueta	0.00	29.11	36.58	52.28	39.00
Libro	0.00	.53	.40	45.17	0.00
Pagina	0.00	.45	2.08	10.11	0.00
URL	0.00	2.01	4.29	5.94	.50
Promedio=	233.95	12.94	48.98	114.33	327.36
Categorización:	Alto	Muy bajo	Bajo	Medio	Muy alto

Para una mejor visualización de los datos, a continuación se presenta la figura 13 que resume la proporción del uso del sistema UADY Virtual de acuerdo a los cinco grupos de profesores detectados.

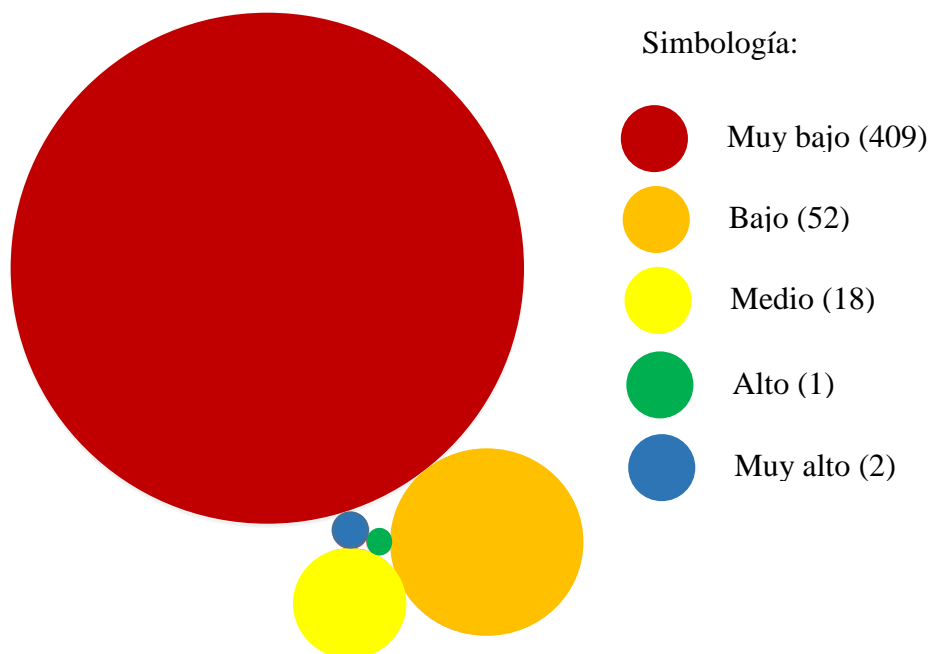


Figura 13. Uso del sistema UADY Virtual de acuerdo a los grupos de profesores.

Para una mejor comprensión y análisis de los resultados, se eliminaron los atributos cuyos valores para los cinco grupos fueran menor a 1, lo cual significaba que el uso de la herramienta era nula. Como resultado se eliminaron los siguientes atributos: Asistencia, Base de Datos, Chat, Cuestionario, Encuesta, Encuesta Predefinida, Glosario, Hot Potatoes y Wiki. Asimismo se eliminaron los decimales sin redondear los valores y se recalcularon los promedios sin que afectase la categorización de los grupos; al final se obtuvo la siguiente tabla:

Tabla 4.

Características de los grupos de profesores. Atributos con un uso mayor a 1.

Variable	Grupo 1 (1)	Grupo 2 (409)	Grupo 3 (52)	Grupo 4 (18)	Grupo 5 (2)
CursoM5	100%	86%	84%	88%	100%
AcreditoM5	100%	76%	82%	88%	100%
Consulta	0	10	28	89	0
Foro	0	2	5	13	0
Lecciones	0	0	0	1	0
SCORM	0	0	0	3	0
Taller	0	0	0	9	0
Tareas	5181	69	852	2182	7289
Archivo	0	14	21	29	0
Carpeta	0	3	5	7	0
Etiqueta	0	29	36	52	39
Libro	0	0	0	45	0
Pagina	0	0	2	10	0
URL	0	2	4	5	0
Promedio=	384	21	80	187	538
Categorización:	Alto	Muy bajo	Bajo	Medio	Muy alto

De la tabla 4 se destacan los siguientes aspectos:

1. La herramienta con mayor uso fue Tareas.

2. Para el grupo con muy alto desempeño (grupo 5), se observa que únicamente tuvieron una herramienta superior (en cuanto a la cantidad de uso) en comparación con los demás grupos.
3. El grupo con la mayor cantidad de herramientas usadas fue el grupo 4, sin embargo, obtuvo el desempeño Medio.

A continuación se presenta la figura 14 que representa visualmente, para cada una de las áreas de conocimiento, el desempeño obtenido de cada profesor. Nota: la nomenclatura de las áreas de conocimiento (campus), es la misma que la presentada en la página 50.

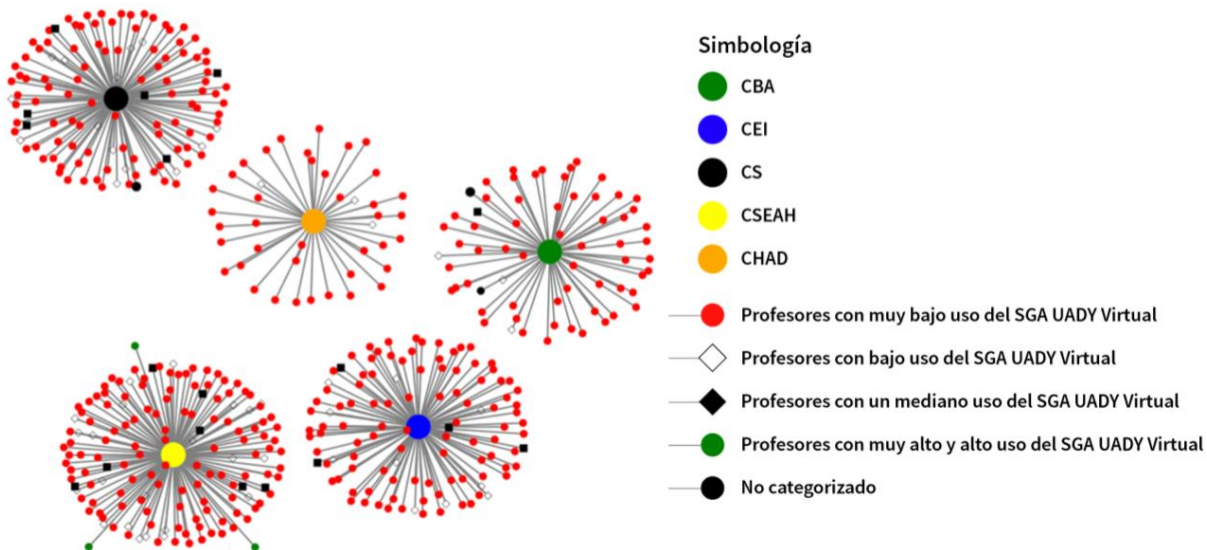


Figura 14. Uso del sistema UADY Virtual de acuerdo a los campus.

A continuación se describe la generación de grupos a través del software WEKA:

El conjunto de datos experimentó algunos ajustes para mejorar los resultados, adicionalmente a las actividades y recursos de UADY virtual se incluyó la variable denominada Clasificación de uso, la cual representa el promedio del desempeño de los

profesores con las herramientas del sistema y cuyos valores se representan en cinco categorías: Muy alto, Alto, Medio, Bajo y Muy bajo.

Posteriormente, el conjunto de datos se transformó al formato ARFF (del inglés, *Attribute-Relation File Format*), el cual es compatible con el software WEKA. Después de experimentar con los algoritmos EM (Expectation-Maximization) Dempster, Laird, Rubin (1977), FilteredClustered (Cartan, 1937) y SimpleKMeans (, 1967), se obtuvieron resultados más eficientes con éste último algoritmo (ver tabla 5).

Tabla 5.

Resultados obtenidos con el algoritmo de agrupamiento SimpleKmeans a través del software WEKA.

Variable	Grupo 1 (2)	Grupo 2 (219)	Grupo 3 (263)
Asistencia	0.50	0.04	0.07
Base de Datos	0.00	0.20	0.16
Chat	0.00	0.02	0.14
Consulta	585.00	11.14	14.18
Cuestionario	2.5	0.50	0.49
Encuesta	0.00	0.22	0.12
Encuesta predefinida	0.00	0.01	0.36
Foro	77.50	2.73	3.56
Glosario	0.00	0.25	0.54
HotPotatoes	0.00	0.01	0.01
Lecciones	0.00	0.01	0.08
SCORM	31.5	0.01	0.03
Taller	89.50	0.14	0.04
Tareas	2476.00	185.57	328.07
Wiki	0.00	0.04	0.15
Archivo	8.00	18.21	13.74
Carpeta	21.00	4.66	2.85
Etiqueta	143.50	34.09	26.90
Libro	406.50	0.65	0.36
Pagina	50.50	0.43	1.06
URL	0.00	1.93	2.77
Clasificación de uso	Medio	Muy bajo	Muy bajo

En la tabla 5, se observa que los elementos mutuamente excluyentes que definen los tres grupos son: Consulta, Tareas (utilizada por la mayoría de los profesores), Archivo, Carpeta y Etiqueta. En cuanto a la variable denominada Clasificación de uso, se reconoce que 482 profesores de los 484 que participaron en el estudio se encuentran clasificados por su desempeño como Muy bajo. En contraste, solamente dos profesores están clasificados como desempeño Medio, debido a que hacen un uso elevado de las actividades Foro, SCORM y Taller, así como de los recursos Libro y Página.

4.2 Técnicas de clasificación

Como resultado de la aplicación del algoritmo J48, se obtuvo el siguiente modelo predictivo representado a través de un diagrama de árbol de decisión:

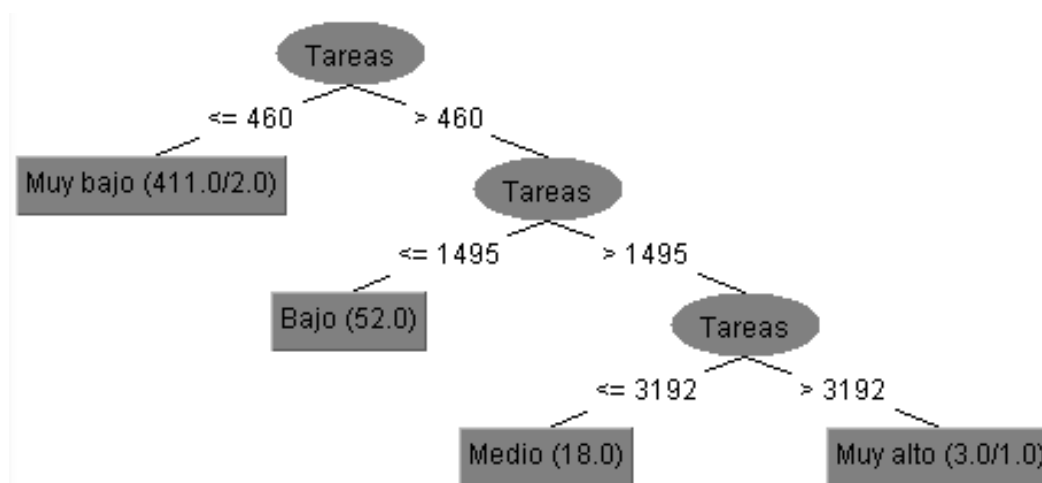


Figura 15. Reglas de clasificación.

En lenguaje natural, el diagrama anterior puede ser interpretado de la siguiente manera:

- Si la actividad tareas se usa por encima de 3192 veces, entonces, la calificación de profesor será muy alto.

- Si la actividad tareas es menor o igual a 3192 y mayor a 1495, entonces, la calificación del profesor será medio.
- Si la actividad tareas es menor o igual a 1495 y mayor a 460, entonces, la calificación del profesor será bajo.
- Finalmente, si la actividad tareas es menor o igual a 460, entonces, la calificación del profesor será muy bajo.

4.3 Técnicas de asociación

Para la obtención de las reglas de asociación, se consideraron las mejores 10 reglas de asociación con las siguientes condiciones: soporte de 0.95, y confianza de .9. Se utilizó el algoritmo A priori. Los resultados se muestran en la tabla 6.

Tabla 6.

Reglas de asociación de las herramientas de UADY Virtual.

#	Reglas de asociación	Interpretación en lenguaje natural
1.	Hot Potatoes = Muy poco ==> Lecciones = Muy poco	Si el profesor usa la herramienta Hot Potatoes muy poco, entonces usará muy poco la herramienta Lecciones.
2.	SCORM = Muy poco ==> Taller = Muy poco	Si el profesor usa la herramienta SCORM muy poco, entonces usará muy poco la herramienta Taller.
3.	SCORM = Muy poco ==> Libro = Muy poco	Si el profesor usa la herramienta SCORM muy poco, entonces usará muy poco la herramienta Libro.
4.	SCORM = Muy poco Libro = Muy poco ==> Taller = Muy poco	Si el profesor usa la herramienta SCORM muy poco y Libro muy poco, entonces usará muy poco la herramienta Taller.

Tabla 6 (continuación)

#	Reglas de asociación	Interpretación en lenguaje natural
5.	SCORM = Muy poco Taller = Muy poco ==> Libro = Muy poco	Si el profesor usa la herramienta SCORM muy poco y Taller muy poco, entonces usará muy poco la herramienta Libro.
6.	SCORM = Muy poco ==> Taller = Muy poco Libro=Muy poco	Si el profesor usa la herramienta SCORM muy poco, entonces usará muy poco las herramientas Taller y Libro.
7.	Hot Potatoes = Muy poco Taller = Muy poco ==> Lecciones = Muy poco	Si el profesor usa la herramienta Hot Potatoes muy poco y Taller muy poco, entonces usará muy poco la herramienta Lecciones.
8.	Hot Potatoes = Muy poco Libro = Muy poco ==> Lecciones = Muy poco	Si el profesor usa la herramienta Hot Potatoes muy poco y Libro muy poco, entonces usará muy poco la herramienta Lecciones.
9.	Lecciones = Muy poco SCORM = Muy poco ==> Taller = Muy poco	Si el profesor usa la herramienta Lecciones muy poco y SCORM muy poco, entonces usará muy poco la herramienta Taller.
10.	Lecciones = Muy poco SCORM = Muy poco ==> Libro = Muy poco	Si el profesor usa la herramienta Lecciones muy poco y SCORM muy poco, entonces usará muy poco la herramienta Libro.

Capítulo 5. Resultados del Análisis Descriptivo e Inferencial

En el presente capítulo se exponen los resultados del análisis descriptivo e inferencial, cabe recordar que los indicadores generados en el análisis descriptivo fueron desarrollados a petición del DIIE, lo que al mismo tiempo da cumplimiento al objetivo específico cuatro, mientras que el análisis inferencial se realizó para dar cumplimiento al objetivo específico cinco.

5.1 Análisis descriptivo

El propósito del análisis descriptivo fue generar indicadores que provean una idea general del uso del sistema UADY Virtual por parte de los profesores de la Universidad. A continuación se muestran los indicadores con sus respectivas figuras y/o tablas.

En la figura 16 se puede observar el porcentaje de profesores de la UADY, que impartieron al menos un curso de nivel de licenciatura y/o posgrado con apoyo del sistema UADY Virtual en el periodo enero – julio 2016.

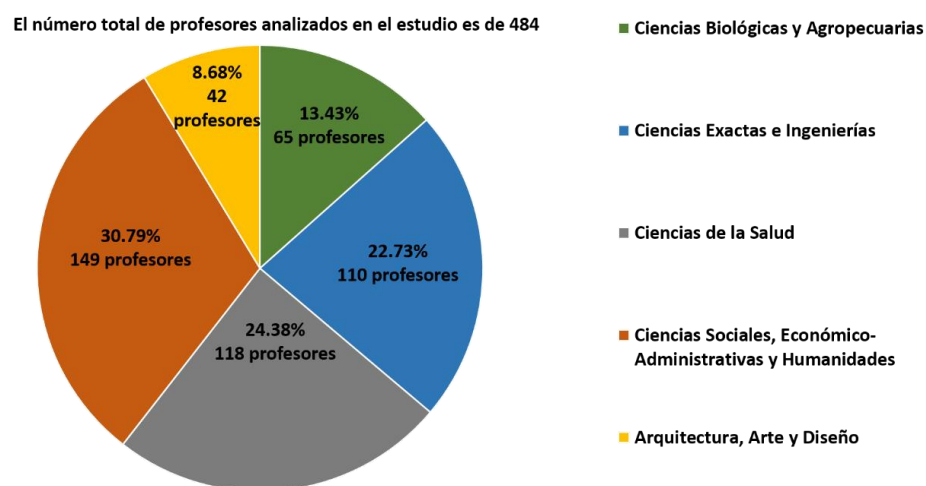


Figura 16. Profesores que impartieron al menos un curso con apoyo del sistema UADY Virtual.

La figura anterior revela el porcentaje y número de profesores por área de conocimiento que usaron el sistema. Se destaca que fue el campus de Ciencias Sociales, Económico Administrativas y Humanidades, el área de conocimiento con la mayor cantidad de profesores activos en el sistema con un 30.79%, mientras que en el campus de Arquitectura, Arte y Diseño, 42 profesores usaron en apoyo a sus clases el sistema, lo que representa un 8.68% de la población total de profesores de la Universidad que emplearon el sistema. El total de profesores que usaron al menos una vez UADY Virtual fueron 484.

Asimismo, la figura 17 muestra el porcentaje de profesores que impartieron al menos un curso en el sistema UADY Virtual y que a su vez acreditaron el Módulo V del Programa Institucional de habilitación del MEFI.

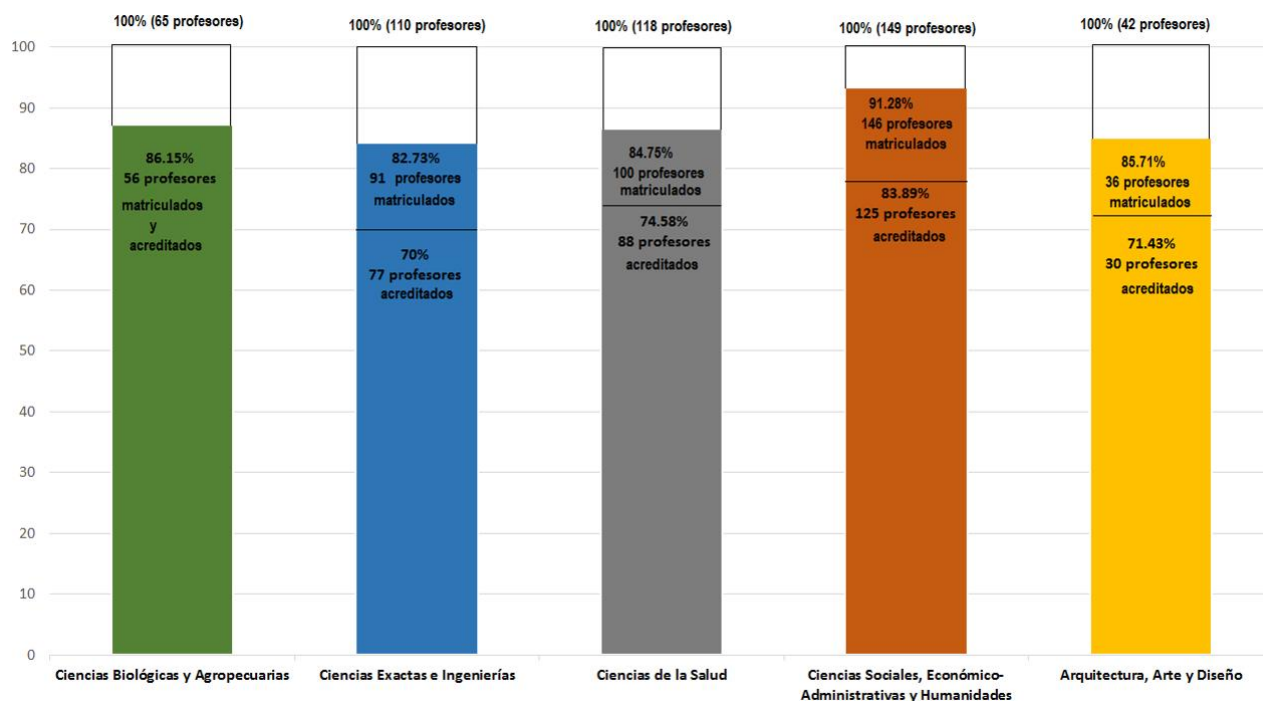


Figura 17. Profesores habilitados por el Módulo 5 del MEFI.

Cabe aclarar que un profesor matriculado es aquél que se ha inscrito al curso de habilitación, pero no ha aprobado el mismo, por otro lado, el profesor acreditado es aquél que ha aprobado satisfactoriamente el curso.

De la figura 17 se puede destacar que en el campus de Ciencias Biológicas y Agropecuarias la cantidad de profesores que cursaron el módulo V (matriculados), es igual al número de profesores que lo acreditaron, es decir, todos los profesores que cursaron el módulo de capacitación, lo aprobaron, sin embargo, no todos los profesores que usaron la plataforma cursaron el módulo.

Por otra parte se observa que el campus con la mayor proporción de profesores matriculados corresponde al campus de Ciencias Sociales, Económico Administrativas y Humanidades, sin embargo, de los 91.28% de profesores matriculados, se acreditaron el 83.89%.

El campus con la mayor diferencia proporcional entre profesores matriculados y profesores acreditados fue el campus de Arquitectura, Arte y Diseño.

A continuación, se muestra por área de conocimiento las actividades y recursos que fueron usadas al menos una vez durante el mismo periodo de análisis (enero – julio 2016).

Para la comprensión de la tabla es importante considerar la siguiente terminología:

- CB: Campus de Ciencias Biológicas y Agropecuarias.
- CE: Campus de Ciencias Exactas e Ingenierías.
- CS: Campus de Ciencias de la Salud.

- CSH: Campus de Ciencias Sociales, Económico-Administrativas y Humanidades.
- CA: Campus de Arquitectura, Arte y Diseño.

Tabla 7.

Actividades y recursos empleados al menos una vez por área de conocimiento.

Actividades	Áreas de Conocimiento				
	CB	CE	CS	CSH	CA
Asistencia	Si	Si	Si	Si	Si
Base de datos	No	Si	Si	Si	Si
Chat	No	Si	Si	Si	Si
Consulta	Si	Si	Si	Si	Si
Cuestionario	Si	Si	Si	Si	Si
Encuesta	Si	Si	Si	Si	No
Encuesta predefinida	No	Si	Si	Si	No
Foro	Si	Si	Si	Si	Si
Glosario	Si	Si	Si	Si	Si
Hot Potatoes	No	No	Si	Si	No
Lecciones	No	Si	Si	Si	Si
SCORM	No	Si	No	Si	No
Taller	No	Si	Si	Si	No
Tareas	Si	Si	Si	Si	Si
Wiki	No	Si	Si	Si	Si
Recursos					
Archivo	Si	Si	Si	Si	Si
Carpeta	Si	Si	Si	Si	Si
Etiqueta	Si	Si	Si	Si	Si
Libro	Si	Si	Si	Si	Si
Página	Si	Si	Si	Si	Si
URL	Si	Si	Si	Si	Si

En la tabla 7, se puede destacar que todas las áreas de conocimiento usaron al menos una vez todos los recursos, a diferencias de las actividades que no fueron usadas en su totalidad, excepto por el área de Ciencias Sociales Económico Administrativas y Humanidades, que sí empleo todas las actividades en al menos una ocasión, mientras que el campus de Biológicas y Agropecuarias usó menos actividades. En complemento de la tabla

anterior, a continuación se muestra el porcentaje de uso de las actividades y recursos del sistema UADY Virtual.

Tabla 8.

Porcentaje de uso de las actividades y recursos del sistema UADY Virtual.

Actividades	Número máximo de accesos	Porcentaje de uso
Asistencia	29	0.02%
Base de datos	87	0.05%
Chat	41	0.02%
Consulta	7341	4.34%
Cuestionario	244	0.14%
Encuesta	84	0.05%
Encuesta predefinida	98	0.06%
Foro	1692	1%
Glosario	196	0.12%
Hot Potatoes	7	0%
Lecciones	25	0.01%
SCORM	74	0.04%
Taller	221	0.13%
Tareas	131876	78.03%
Wiki	51	0.03%
Recursos		
Archivo	7619	4.51%
Carpeta	1814	1.07%
Etiqueta	14830	8.77%
Libro	1052	0.62%
Página	476	0.28%
URL	1155	0.68%
Total	169012	100%

En la tabla 8, se puede observar el elevado porcentaje de uso de la actividad Tareas con un 78.03%, adicionalmente, únicamente las actividades Consulta y Foro superan el 1% de uso; respecto a los recursos, sólo Archivo, Carpeta y Etiqueta superan el 1% de uso. Todas las demás actividades y recursos están por debajo del 1%.

Para fines posteriores de análisis, se consideraron las actividades y recursos de alto uso, aquellas que superaron el 1% de uso, mientras que las actividades y recursos que no superaron dicho porcentaje, fueron consideradas de bajo uso.

A continuación se presenta la figura 18 que muestra la distribución en porcentaje del uso de las actividades y recursos derivados de la tabla anterior.

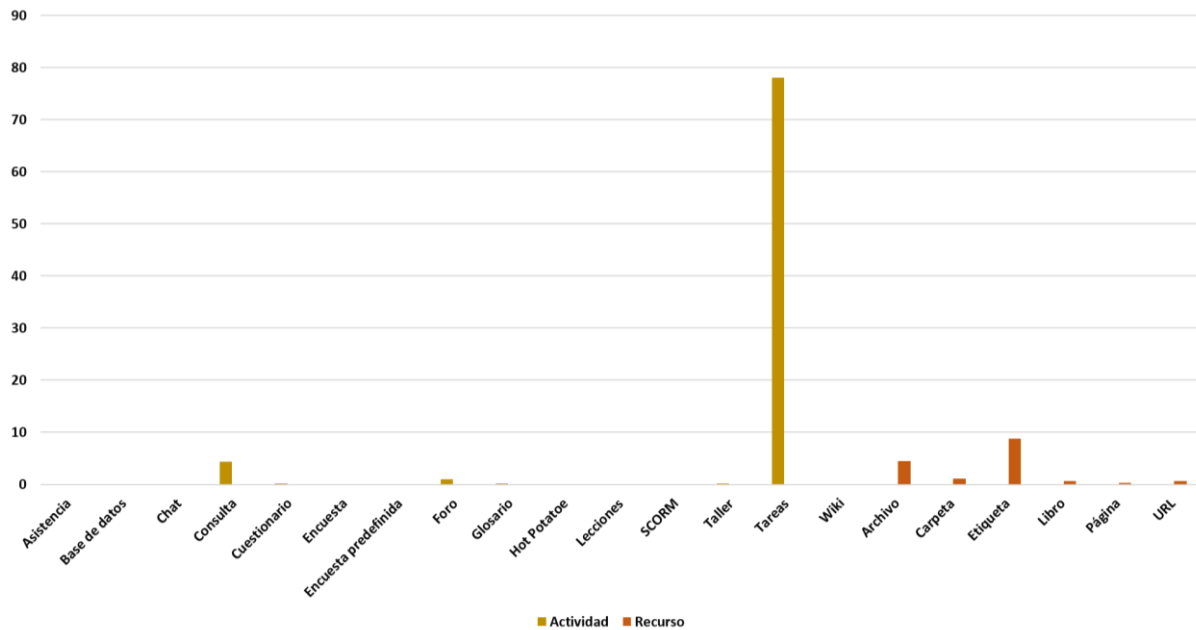


Figura 18. Porcentaje de uso de las actividades y recursos en UADY Virtual.

La siguiente figura (figura 19), muestra el uso de cada una de las herramientas de UADY Virtual usadas durante el periodo enero – julio del 2016 dividido en cada una de las áreas de conocimiento.

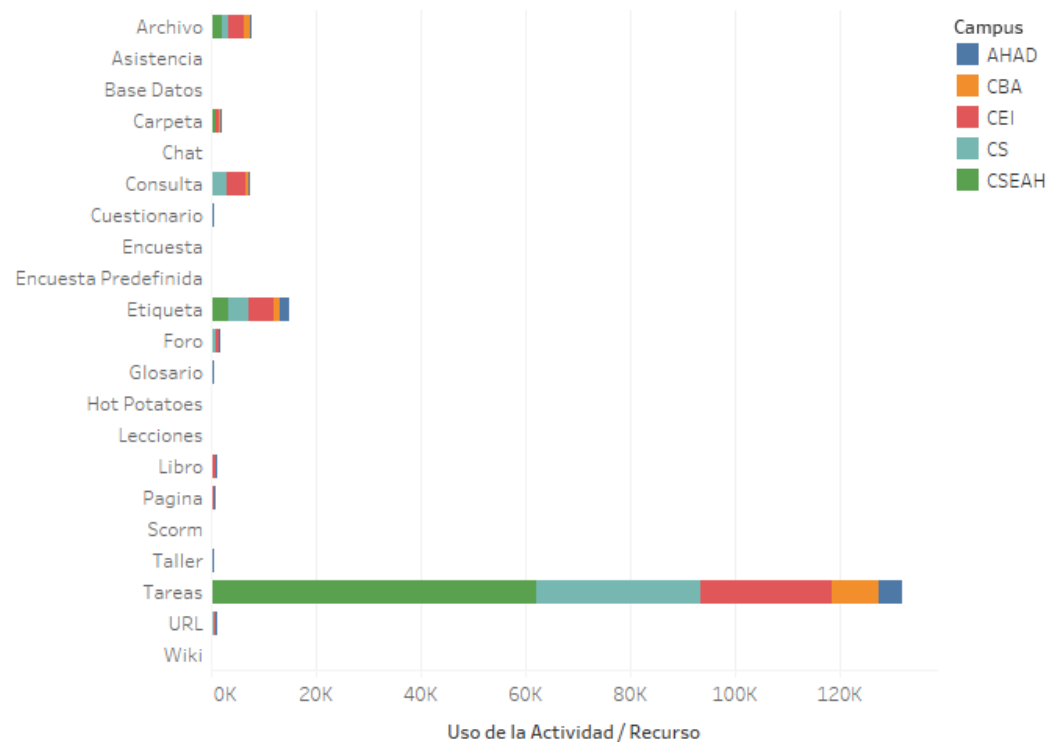


Figura 19. Porcentaje de uso de las actividades y recursos en UADY Virtual dividido por áreas de conocimiento.

De la figura anterior se destaca que, en la actividad tareas, el CSEAH tuvo mayor peso en el uso de la misma, sin embargo, no significa que fue el campus que más la uso, ya que el CSEAH posee más programas y profesores alineados al MEFI y en consecuencia, mayor cantidad de registros en UADY Virtual.

Finalmente se exponen las figuras 20, 21, 22 y 23, que muestran respectivamente, para cada área de conocimiento, el porcentaje de las actividades de uso alto (1% o más), de uso bajo (menos del 1%), el porcentaje de los recursos de uso alto (1% o más) y de uso bajo (menos del 1%).

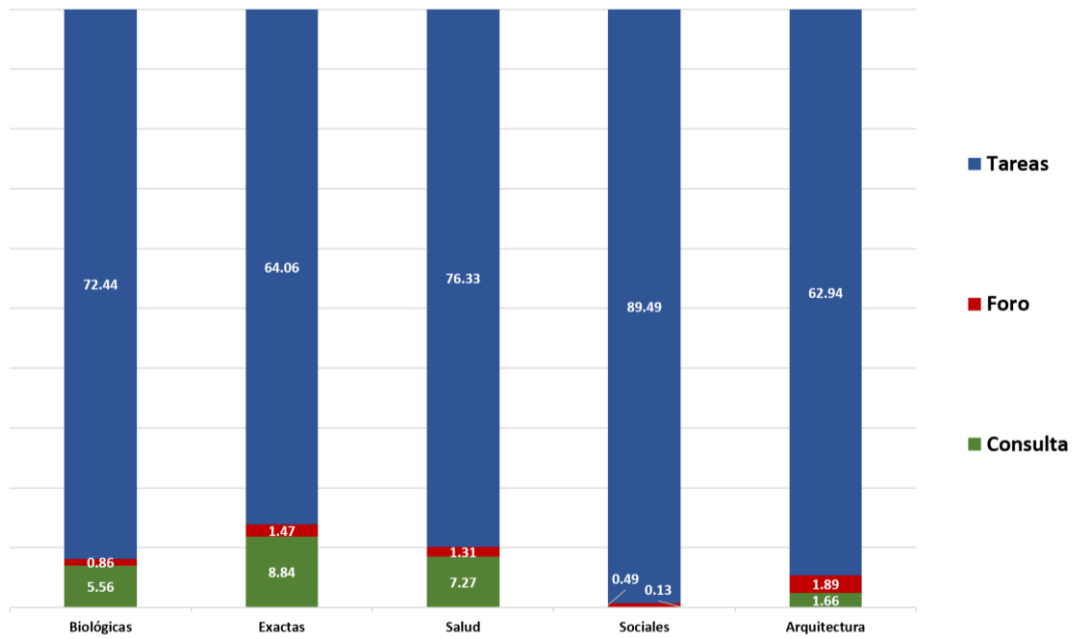


Figura 20. Porcentaje de las actividades de uso alto.

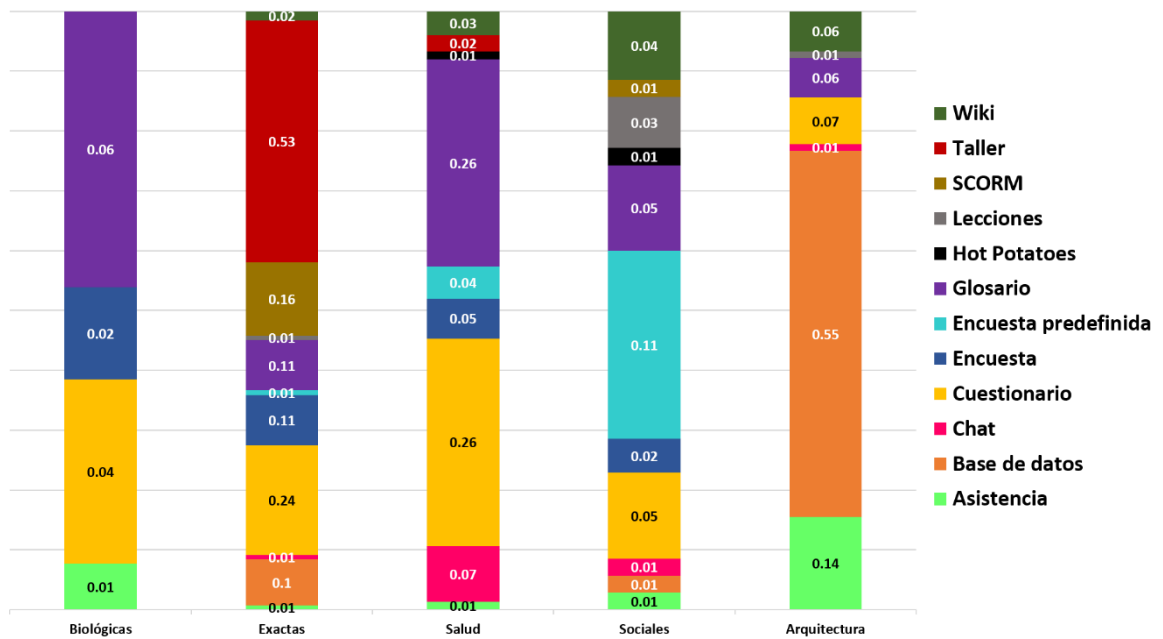


Figura 21. Porcentaje de las actividades de uso bajo.

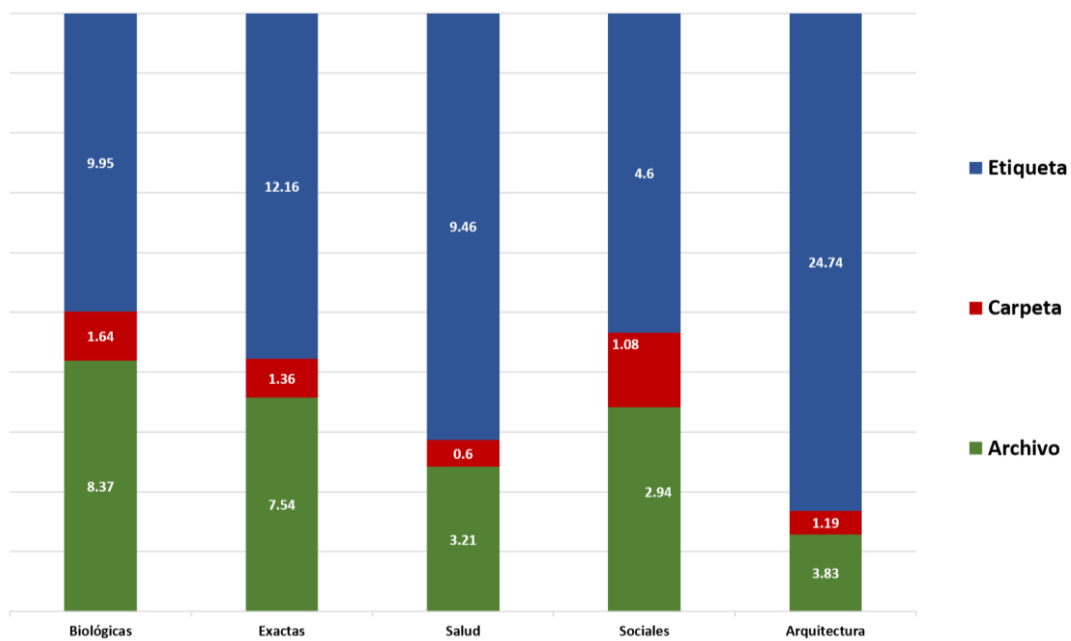


Figura 22. Porcentaje de los recursos de uso alto.

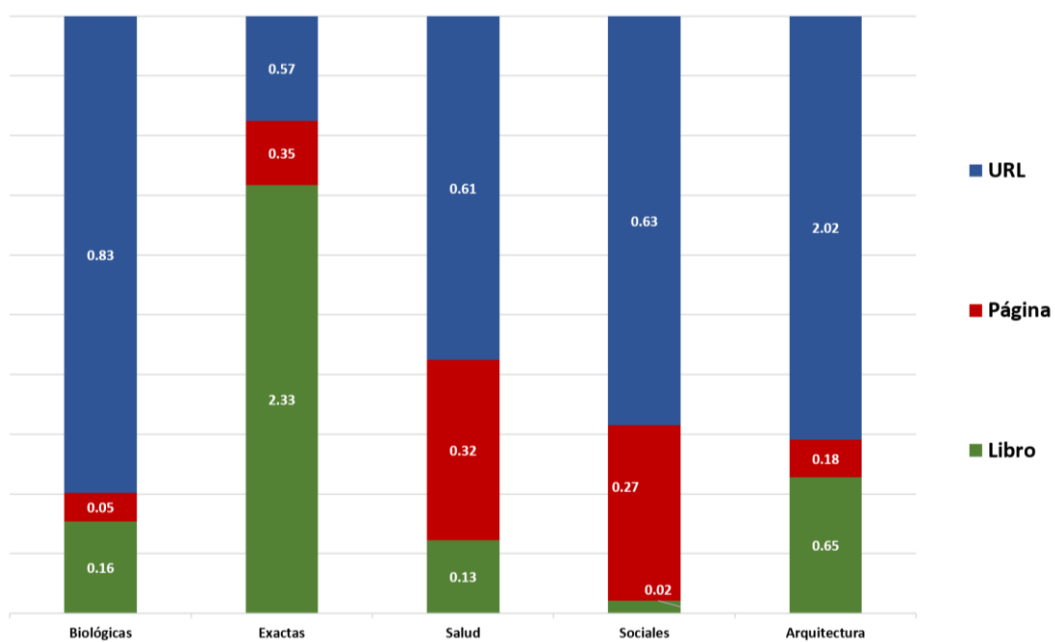


Figura 23. Porcentaje de los recursos de uso bajo.

Finalmente, la figura 24 muestra la distribución del uso de las herramientas de acuerdo a las áreas de conocimiento.

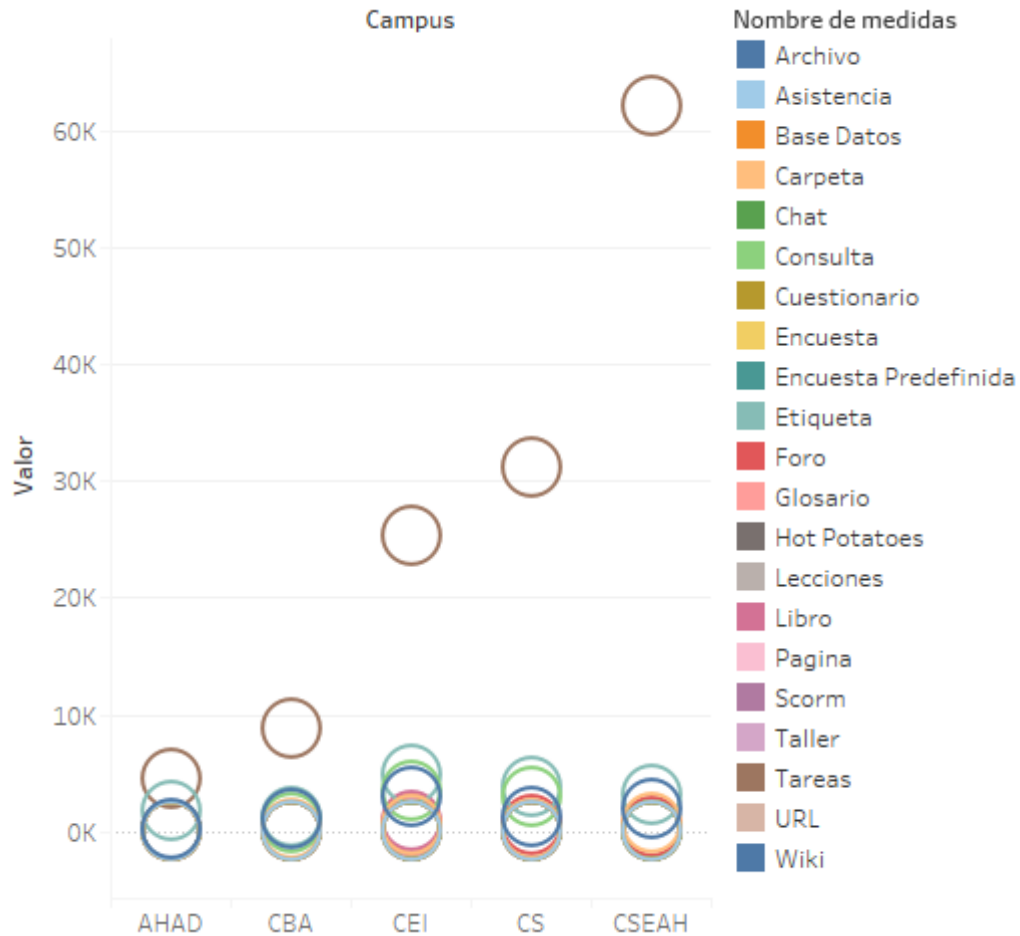


Figura 24. Distribución de uso de las herramientas de UADY Virtual para cada área de conocimiento.

De la figura 24 se destaca que en las cinco áreas de conocimiento la herramienta más usada fue la actividad tareas. En cuanto al uso diversificado de las herramientas destacan CEI y CS, es decir, dichas áreas de conocimiento emplearon más herramientas en comparación con los demás Campus. El área de conocimiento con menor uso de herramientas fue CBA. Las comparaciones antes señaladas serán comprobadas a través de análisis inferenciales.

5.2 Análisis inferencial

El propósito de la aplicación del análisis inferencial, fue determinar si existieron diferencias significativas en el uso de UADY Virtual de acuerdo al área de conocimiento, para ello, en primer lugar se realizó la prueba de normalidad de Kolmogorov – Smirnov (Lilliefors, 1967) ya que la población (profesores de la UADY) era mayor a 50 sujetos. El objetivo de la aplicación de la prueba de normalidad fue determinar si la distribución de cada una de las variables numéricas (herramientas de UADY Virtual) tenía una distribución normal o anormal (no paramétrica).

Para la prueba de normalidad se analizaron tanto cada una de las variables (herramientas de UADY Virtual) como una variable global (la cual consideraba todas las variables al mismo tiempo), en todos los análisis se consideró como hipótesis nula que, la distribución de la variable a analizar es igual a la distribución normal, mientras que la hipótesis alterna consistió en que la distribución de la variable a analizar era diferente de la distribución normal. Se consideró un nivel de significancia de 5% (.05) para la región de rechazo. Los análisis se desarrollaron con el software SPSS.

En ninguno de los casos (análisis individual de las variables y análisis de los datos global) el nivel de significancia fue mayor a .05, es decir, en todos los análisis $P < .05$, por lo tanto se rechazó (en todos los casos) la hipótesis nula, es decir, en todos los casos la distribución de la variable a analizar fue distinta de la distribución normal, comprobando que los datos tenían una distribución asimétrica, es decir no paramétrica o anormal.

A continuación se muestra la figura 25 que comprueba que la distribución de la variable global (total de herramientas) es distinta a la distribución normal.

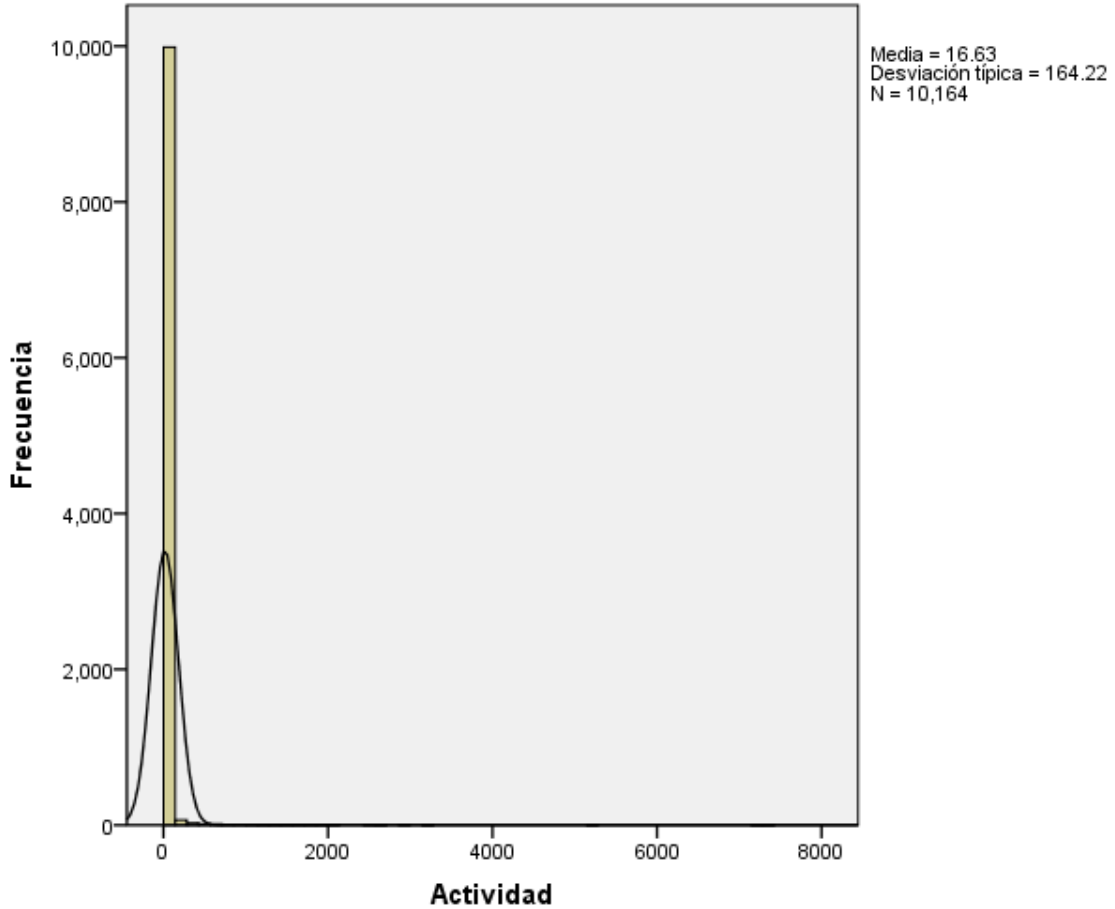


Figura 25. Prueba de normalidad para la variable global (total de herramientas usadas).

A continuación se presenta la tabla 9 que resume los resultados de normalidad aplicado a cada una de las herramientas.

Tabla 9.

Prueba de Kolmogorov – Smirnov para cada una de las herramientas.

Actividades	Z de la prueba	Significancia
Asistencia	11.48	0.0
Base de datos	11.23	0.0
Chat	11.56	0.0
Consulta	10.92	0.0
Cuestionario	9.72	0.0
Encuesta	11.06	0.0
Encuesta predefinida	9.24	0.0
Foro	7.92	0.0

Tabla 9 (continuación)

Actividades	Z de la prueba	Significancia
Glosario	10.77	0.0
Hot Potatoes	11.54	0.0
Lecciones	11.31	0.0
SCORM	11.37	0.0
Taller	11.16	0.0
Tareas	7.65	0.0
Wiki	11.46	0.0
Recursos		
Archivo	6.74	0.0
Carpeta	7.75	0.0
Etiqueta	5.54	0.0
Libro	10.45	0.0
Página	10.24	0.0
URL	8.48	0.0

Debido a que ninguna de las variables (herramientas), tenían una distribución normal, se decidió aplicar una prueba de hipótesis para datos no paramétricos. Se utilizó la prueba de Kruskal-Wallis (Daniel, 1990) para comparar el uso de cada herramienta entre áreas de conocimiento, en caso de diferencias significativas se aplicó la comparación múltiple de Dunn (Daniel, 1990). Las pruebas estadísticas se consideraron significativas cuando $P < 0.05$ y se utilizaron los paquetes estadísticos STATGRAPHICS Centurion XVII v. 17.0.16 (Statpoint, 2014) y el SPSS 22 (IBM SPSS, 2013).

Cabe destacar que no se distinguieron cuántos cursos impartió cada profesor ($n=484$), por lo que cada valor es una medida de frecuencia global para cada herramienta sin importar el número de cursos impartidos por profesor. Para lograr lo anterior se creó la base de datos UADY Virtual V3.2, en donde se recodificaron los valores de los atributos herramientas, pasando de un valor numérico a un 1, si la herramienta (para cada profesor) tenía algún registro, y a un 0, si la herramienta (para cada profesor) no tenía ningún registro.

Para la comprensión de los resultados, es importante considerar la siguiente nomenclatura (ver tabla 10).

Tabla 10.

Nomenclatura Áreas de Conocimiento (AC).

Sigla	Área de Conocimiento (AC)
AHAD	Arquitectura, Hábitat, Arte y Diseño.
CBA	Ciencias Biológicas y Agropecuarias.
CEI	Ciencias Exactas e Ingenierías.
CS	Ciencias de la Salud
CSEAH	Ciencias Sociales Económico Administrativas y Humanidades

5.2.1 Herramientas con diferencias significativas de acuerdo al campus

A continuación se muestran las herramientas en donde hubieron diferencias significativas ($P < .05$) de acuerdo al área de conocimiento, al mismo tiempo se redacta (derivado de las comparaciones múltiples) las AC con mayor y menor uso para la herramienta en cuestión.

1. Base de Datos ($H=13.1904$, $P=0.0042$, $gl=3$), resultó significativamente mayor en CEI que en CS y CSEAH.
2. Chat ($H=14.0358$, $P=0.0029$, $gl=3$), siendo mayor en CS que en las demás AC.
3. Consulta ($H=68.2897$, $P < 0.0001$, $gl=4$), tanto para CS como para CEI resultó mayor que en AHAD, CSEAH y CBA.
4. Cuestionario ($H=34.5545$, $P < 0.0001$, $gl=4$), resultó significativamente mayor en CEI que en AHAD, CBA y CSEAH, también fue mayor en CS que en AHAD y CBA, sin embargo, no difirió entre CEI y CS.

5. Encuesta ($H=15.023$, $P=0.0018$, $gl=3$), siendo mayor en CEI que en CBA y CSEAH. Cabe señalar que en el AC AHAD no se registró uso de la herramienta Encuesta.
6. Encuesta Predefinida ($H=11.4977$, $P=0.0032$, $gl=2$), en CS se obtuvo significativamente mayor uso de la herramienta que en CEI y CSEAH. Cabe señalar que en las AC AHAD y CBA no se registró uso de la herramienta Encuesta Predefinida.
7. Foro ($H=29.1626$, $P<0.0001$, $gl=4$), tanto en CS como en CEI resultó mayor que en los otros tres campus, además en CS y CEI.
8. Glosario ($H=10.2648$, $P=0.0362$, $gl=4$), fue significativamente mayor en CS que en CBA, AHAD, y CSEAH, sin embargo, no difirió entre CS y CEI.
9. Tareas ($H=25.4753$, $P<0.0001$, $gl=4$), tanto en CSEAH como en CS resultó mayor que en CBA y AHAD.
10. Archivo ($H=26.4592$, $P<0.0001$, $gl=4$), en CEI se obtuvo significativamente mayor uso que en CSEAH, AHAD y CS.
11. Carpeta ($H=15.5562$, $P=0.0037$, $gl=4$), en CEI resultó significativamente mayor que en AHAD y CS.
12. Etiqueta ($H=23.6580$, $P<0.0001$, $gl=4$), en CEI resultó significativamente mayor que en CSEAH y CBA.
13. Libro ($H=23.0499$, $P=0.0001$, $gl=4$), en CEI se obtuvo significativamente mayor uso que en CSEAH, CBA y CS.

14. Página ($H=16.3907$, $P=0.0025$, $gl=4$), en CEI resultó significativamente mayor que en CBA y CSEAH.

5.2.2 Diferencia del uso de UADY Virtual de acuerdo al campus

Finalmente, se utilizó nuevamente la prueba no paramétrica de Kruskal-Wallis para comparar el Total de Herramientas Usadas (THU) entre áreas de conocimiento, en caso de diferencias significativas se aplicó la comparación múltiple de Dunn. La variable THU puede describirse de la siguiente manera:

$$THU = \sum_{i=1}^{21} H_i \quad (2)$$

Donde $H_i=1$ si el profesor usó la herramienta i -ésima y $H_i=0$ si el profesor no usó la herramienta i -ésima. Así, para cada profesor se tiene $0 \leq \sum_{i=1}^{21} H_i \leq 21$, es decir, $0 \leq THU \leq 21$

A continuación el resultado: El THU entre AC difirió significativamente ($H=42.4941$, $P<0.0001$, $gl=4$), resultando mayor en: 1) CEI que en CSEAH, AHAD y CBA, y 2) CS que en CSEAH y AHAD. Sin embargo no difirió entre CEI y CS (ver Tabla 11).

Tabla 11.

Total de herramientas usadas entre áreas de conocimiento.

Campus	n	Promedio	DE	Rango Promedio	Mediana	Mínimo	Máximo
CSEAH	149	3.86577	2.51927	202.19 a	4	1	12
AHAD	42	3.85714	2.32244	202.62 a	3	1	11
CBA	65	4.01538	1.90798	220.06 a b	4	1	10
CS	118	5.01695	2.94967	261.26 b c	5	1	15
CEI	110	5.88182	2.8308	305.46 c	5.5	1	13

Nota: Los rangos promedio con distinta letra difieren ($P<0.05$), prueba de Dunn.

Capítulo 6. Conclusiones

En este capítulo se describe el análisis de la consecución de los cinco objetivos específicos que se plantearon en la introducción. Posteriormente, se presenta la discusión de los resultados obtenidos. Adicionalmente, se muestran recomendaciones y trabajo a futuro. Finalmente, se exponen las referencias de los trabajos divulgados de esta investigación.

6.1 Análisis de la consecución de los objetivos

El planteamiento del problema que desencadenó el desarrollo de la investigación fue el desconocimiento por parte de la UADY respecto al uso que le dan los profesores a UADY Virtual. Derivado de la problemática planteada, se conformó la siguiente pregunta de investigación: ¿Existen necesidades de formación que pudiesen estar requiriendo los profesores de la UADY de acuerdo a la utilización del sistema de gestión del aprendizaje UADY Virtual?

Para dar respuesta a la pregunta de investigación, se planteó el siguiente objetivo general: determinar si existen necesidades de formación en el uso de las herramientas del sistema de gestión de aprendizaje UADY Virtual. Enseguida, con el propósito de dar cumplimiento al objetivo general, se establecieron cinco objetivos específicos.

A continuación se exponen, para cada uno de los objetivos específicos, las conclusiones y reflexiones finales, componentes que permiten verificar el cumplimiento de los mismos.

1. Crear perfiles de usuario de profesores de acuerdo a su desempeño con UADY

Virtual

Para la generación de los perfiles, se empleó la técnica de minería de datos agrupamiento, con el propósito de calificar el desempeño de los profesores de acuerdo a la magnitud del uso del sistema por parte de los mismos. Las calificaciones se etiquetaron de la siguiente manera: muy bajo, bajo, medio, alto y muy alto.

De los resultados obtenidos con el software SPSS se observa que 409 profesores (de 484) pertenecen al grupo muy bajo, lo que significa que, al ser el 84.5% de la población total perteneciente al grupo más bajo, se corrobora la existencia de necesidades docentes en materia de capacitación, motivación e incentivación.

En cuanto a los resultados obtenidos con el software WEKA se reconoció que 482 profesores de los 484 que participaron en el estudio se encuentran clasificados por su desempeño como Muy bajo. En contraste, solamente dos profesores están clasificados como desempeño Medio, debido a que hacen un uso elevado de las actividades Foro, SCORM y Taller, así como de los recursos Libro y Página.

2. Construir un modelo de conocimiento que permita predecir el desempeño del profesor en función de la cantidad de veces que emplea ciertas actividades y/o recursos del sistema.

Derivado de las técnicas de clasificación de la minería de datos, se obtuvieron diversas reglas de conocimiento. De las cuales, se identificaron que cuatro de ellas están orientadas a la predicción del desempeño del profesor basadas en el uso de la actividad tareas. Cabe destacar que el modelo predictivo por el momento no cuenta con suficiente

información debido al uso desproporcionado de las actividades y recursos con los que cuenta UADY virtual. Esta área se reconoce como un área de oportunidad para futuras investigaciones.

3. Construir reglas de asociación de las relaciones entre las herramientas de UADY Virtual.

Las reglas de asociación identifican que las correlaciones más fuertes entre las herramientas del sistema se conforman por las siguientes: *hot popatoes*, lecciones, SCORM, taller y libro.

Debido a los resultados y las relaciones identificadas, se concluye que son dichas herramientas las que requieren mayor importancia para ser atendidas respecto a capacitación, motivación y/o incentivación.

Finalmente, se concluye esta sección tomando en cuenta que al ser alcanzados los objetivos específicos, puede considerarse que el objetivo general de la investigación se ha cumplido, ya que se ha determinado que si existieron necesidades de formación docente en el uso de las herramientas del sistema de gestión de aprendizaje UADY Virtual; con ello, asimismo se dio respuesta a la pregunta de investigación.

De igual forma, queda demostrada la veracidad de la hipótesis presentada en el primer capítulo, ya que se dio cumplimiento al objetivo general de la investigación; por lo tanto, puede afirmarse que los profesores de la Universidad Autónoma de Yucatán requieren formación en el uso las herramientas del sistema de gestión de aprendizaje UADY Virtual.

4. Calcular las frecuencias y porcentajes de las actividades y recursos usados por los profesores de la UADY y presentarlas de acuerdo a las cinco áreas de conocimiento.

La base de datos que sirvió como fuente para el análisis de datos, estuvo conformada por la actividad de 484 profesores de las cinco áreas de conocimiento con los que cuenta la UADY en el periodo enero – julio del año 2016. La condición que se consideró es que los profesores utilizaran este sistema como una herramienta de apoyo a sus clases presenciales en los niveles de licenciatura y posgrado.

De los resultados se observa que de las actividades y recursos del sistema, fue la actividad tareas la que registró mayor uso con un 78.03%; la desviación estándar, cuyo valor fue de 20.06, puede ser considerada elevada, lo que significa que la distribución del uso de las actividades y recursos es desproporcionado, es decir, los profesores emplean primordialmente la actividad tareas, mientras que las demás actividades y recursos casi no son usados.

De lo anterior se puede concluir que el uso del sistema por parte de los profesores se asemeja a un repositorio de tareas, en otras palabras, un espacio en donde los estudiantes, a través del mismo, distribuyen y comparten sus actividades de aprendizaje (tareas).

Sería congruente aseverar que el sistema UADY Virtual tiene diversas áreas de oportunidad, ya que contiene tanto herramientas colaborativas (wiki, glosario, foro) como de diagnóstico (encuesta, encuesta predefinida) que no se están utilizando.

Por otra parte, la adopción de los estudiantes a las actividades y recursos con los que cuenta UADY Virtual depende mayormente de la forme como lo utilice el profesor. Decir que el 78% del uso del sistema radica en el uso de una sola herramienta, en este caso,

tareas, significa que existe una enorme brecha entre lo que se espera por parte del profesor (respecto al uso eficaz de las herramientas) y su desempeño real.

Por todo lo anterior, se reconoce que existen oportunidades en cuanto a formar, motivar e incentivar a los profesores, de tal forma que, no sólo se eleve la frecuencia del uso del sistema, sino que además se utilicen las herramientas de manera más distribuida.

5. Determinar si existen diferencias significativas en el uso de las actividades y recursos del sistema por parte de los profesores de la UADY en función de las cinco áreas de conocimiento.

Se efectuó un análisis de varianza de un solo factor para comparar el uso de UADY Virtual entre las cinco áreas de conocimiento de la UADY; para ello se planteó la siguiente hipótesis: ¿existe diferencia significativa en el uso del sistema UADY Virtual de acuerdo al área de conocimiento?

Derivado del análisis de varianza de un factor, se determinó que no existieron diferencias significativas en el uso del sistema UADY Virtual de acuerdo al área de conocimiento, lo que significa que la manera en la que interactúan los profesores con el sistema es igual para todas las áreas de conocimiento. Por lo tanto, se concluye que la afirmación del objetivo específico anterior, es consistente para las cinco áreas de conocimiento, en concreto, las cinco áreas de conocimiento usan como principal herramienta la actividad tareas.

6.2 Discusión

Los resultados fueron consistentes con la investigación efectuada por Romero (2010), en cuanto a que en el manejo del tipo de dato (numérico o nominal) influyen

considerablemente los resultados de la aplicación de los algoritmos, tanto de agrupamiento como de clasificación, asimismo, los algoritmos de clasificación mejoran sus porcentajes predictivos si se usa en totalidad la base de datos, es decir, sin filtro alguno. Respecto a las diferencias, de igual manera, considerando la investigación previamente mencionada, se obtuvo un porcentaje de efectividad del modelo superior al 90%, lo que significa que el modelo clasificó con mejor precisión las actividades de los profesores.

Cabe destacar que 6 de las 21 herramientas registraron un uso superior al 1% (tareas, consulta, foro, archivo, carpeta y etiqueta) de las cuales, archivo, carpeta y etiqueta, es decir, el 50% de las que tuvieron un uso superior, son herramientas cuyo propósito radican en la construcción de la interfaz del sitio web y para la navegación del mismo, por consiguiente, se podría afirmar que sólo las herramientas consulta, tareas y foro tienen un impacto directo sobre el aprendizaje del estudiante.

Los planes a futuro de la UADY plantean retos sin precedentes, la próxima oferta académica de nivel superior en modalidad virtual, requerirá, no sólo mayor capacitación a los profesores respecto al uso de UADY Virtual, sino la aplicación de dichas herramientas en un nuevo enfoque tecno – pedagógico; además de la generación de contenido digital propio.

6.3 Recomendaciones y trabajo futuro

La investigación estable un marco de referencia en cuanto a la metodología para la obtención y procesamiento de datos provenientes de la base de datos de UADY Virtual, por lo tanto, se recomiendan las siguientes acciones concretas:

- Impartición de cursos personalizados tomando en cuenta el área de conocimiento y el desempeño de los profesores, donde se promuevan aquellas actividades y recursos que no reportan actividad.
- Creación de un canal de videos por parte de la institución donde se promuevan aquellas actividades y recursos que reporten un uso nulo en todas las áreas de conocimiento.
- Identificación de profesores con un desempeño por encima de la media para impartirles un curso avanzado, con el objetivo de que ellos se conviertan en instructores de esos cursos de formación personalizados.
- Desarrollo de tutores inteligentes que ayuden a los profesores en el diseño de sus cursos independientemente del nivel de desempeño que tengan.
- Implementación de herramientas de visuales para los administradores de UADY Virtual que permitan obtener indicadores en tiempo real sobre el uso de este sistema.
- Replicar cada semestre este experimento, a fin de medir la evolución del uso del sistema, de igual manera se recomienda un análisis más profundo de los datos, el análisis puede desarrollarse de acuerdo a los siguientes niveles: facultad, programa educativo y asignatura.
- Incentivar investigaciones relacionadas con el análisis del desempeño de los usuarios de UADY Virtual.

Finalmente, se destaca que esta investigación es de mucha utilidad, debido a que la UADY planea crear 45 licenciaturas en línea en los próximos 10 años. Prueba de ello, es que la primera Licenciatura en Educación en modalidad virtual se ofertará en agosto del año 2018.

6.4 Divulgación de resultados

Con el propósito de divulgar los distintos resultados obtenidos de la investigación, se han generado varias publicaciones, las cuales se presentan las referencias:

- Camacho, P., Zapata, A. Menéndez, V.H., Canto, P.J. (2016). Detección de necesidades de formación de profesores en el uso de un sistema de gestión del aprendizaje a través de minería de datos. En Prieto, M., Pech, S. *La tecnología como instrumento para potenciar el aprendizaje*, pp. 475-480. ISBN: 978-1533431110.
- Camacho, P., Zapata, A. Menéndez, V.H., Canto, P.J. (2017). Análisis sobre el desempeño de los profesores en el Sistema de Gestión del Aprendizaje UADY Virtual. En Prieto, M., Pech, S., Zapata, A. *Tecnología y Aprendizaje- Avances en el mundo académico hispano*, pp. 368-375. ISBN: 978-84-697-2772-0.
- Artículo actualmente en revisión en la revista Siglo XXI editada por la Facultad de Educación de la Universidad Nacional de Educación a Distancia de Madrid, España (Indexada en el JCR con un factor de impacto de 1.094).

Referencias

- Abdullateef, B. N., Elias, N., Mohamed, H., Zaidan, A., & Zaidan, B. (2015). Study on Open Source Learning Management Systems: a Survey, Profile, and Taxonomy. *Journal of Theoretical & Applied Information Technology*, 82(1), 93-105.
- Agrawal, R., Imieliński, T., Swami, A. (1993). Mining association rules between sets of items in large databases. In *Acm sigmod record*, 22(2), 207-216.
- Ahmad, J., Gea, M., Paderewski, P., y Gutierrez, F. (2005). A comparison and evaluation Of Open source learning managment system. IADIS International Conference- Applied Computing 2005. Recuperado el 24 de agosto del 2015 de:
<https://MOODLE.org/pluginfile.php/554/.../Jamil1.pdf>
- Ahmad, N. B., Ishak, M. K., Alias, U. F., & Mohamad, N. (2015). An Approach for E-Learning Data Analytics using SOM Clustering. *International Journal of Advances in Soft Computing & Its Applications*, 7(3).
- Alonso, F., López, G., Manrique, D., & Viñes, J. M. (2005). An Instructional Model for Web-Based e-Learning Education with a Blended Learning Process Approach. *British Journal of Educational Technology*, 36(2), 217-235.
- Alavi, M., Leidne, D. (2001). Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *Management Information Systems Research Center, University of Minnesota*, 25(1), 107–136.
- Alcalá-Fdez, J., Sánchez, L., García, S., Jesus, M.J., Ventura, S., Garrell, J.M., Otero, J., et al. (2008). KEEL: a software tool to assess evolutionary algorithms for data mining problems. *Soft Computing*, 13(3), 307–318. doi:10.1007/s00500-008-0323-y

- Anderson, R., & Becker, H. (2001). Report #8 School Investments in Instructional Technology. Recuperado de:
http://www.crito.uci.edu/tlc/findings/report_8/REPORT_8.PDF
- Apereo Foundation. (2016). Sakai. Recuperado de: <https://sakaiproject.org/>
- Banco Mundial (2013). About the World Bank Institute. Recuperado de:
<http://wbi.worldbank.org/wbi/content/acerca-del-instituto-del-banco-mundial>.
- Blackboard (2016). Recuperado de: <http://www.blackboard.com/about-us/index.aspx>
- Boiko, B. (2005). Content management bible. John Wiley & Sons.
- Briceño, M, Poot, P. (2015). Uso de un sistema de gestión del aprendizaje desde la perspectiva de los usuarios. En Prieto, M. E, Pech, S., García, J, De León, T. (Eds.) Aportaciones en el uso de las tecnologías para el aprendizaje, 487-492.
- Cabrero, J. (2007). *Nuevas Tecnologías Aplicadas a la Educación*. McGraw-Hill.
- Cartan, H. (1937). Theorides filters C.R. Acad.Sci.Paris, 205, 595-598
- Cemile, F. (2008). Intelligent learning system for online learning. *International Journal of Hybrid Intelligent Systems* , 129-141.
- CEPAL. División de Desarrollo Social (2014). Los derechos de la infancia en la era de internet: América Latina y las nuevas tecnologías. Recuperado de:
<http://www.cepal.org/es/publicaciones/37049-los-derechos-de-la-infancia-en-la-era-de-internet-america-latina-y-las-nuevas>
- Chen, H. (2001). *Knowledge Management Systems: A text Mining perspective*. University of Arizona, Tucson, Arizona. Knowledge Computing Corporation.

Chen, A. (2006). *The Technology and Application of Data Mining. Beijing: Science Press.*

Clark, David James. (2002). Big Bang or Steady Evolution?. Recuperado de:

http://www.learningtechnologies.co.uk/magazine/article_full.cfm?articleid=6&issueid=7§ion=1

Coenen, F. (2004). Data mining: past, present and future. Extraído de:

http://cgi.csc.liv.ac.uk/~frans/PostScriptFiles/kerDataMining_2010-8-19.pdf

CONACYT (2014). Programa especial de ciencia, tecnología e innovación 2014-2018.

Recuperado de: http://www.conacyt.mx/siicyt/images/PECiTI-2014_2018.pdf

Cooke, Michael. (2004). Clomedia: The Evolution of e-Learning. Recuperado de:

http://www.clomedia.com/content/templates/clo_webonly.asp?articleid=571&zoneid=78

Daniel, W.W. (1990). *Applied Nonparametric Statistics. 2a Ed. Duxbury. Pacific Grove CA, USA.*

Dempster, A.P, Laird, N.M, Rubin, D. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B*

(Methodological), 39(1), 1-38.

Dillenbourg, P. (2000). Virtual Learning Environments. *Learning in the New Millennium:*

Building New Education Strategies for Schools.

Downes, S. (2005). Feature: E-learning 2.0. *eLearn Magazine*, 10.

Dougiamas, M., Taylor, P.C. (2003). MOODLE : Using Learning Communities to Create an Open Source Course Management System. Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications (pp. 171–178.).

DBMiner Technology Inc. (2003). DBMiner. Recuperado el 17 de mayo de <https://cs.wmich.edu/~yang/teach/cs595/Assignment1/dbminer.htm>

DreamBox (2016). DreamBox Learning. Recuperado el 02 de mayo 2016 de:

<http://www.dreambox.com/>.

Drucker, P. (1992). The Age of Discontinuity: Guidelines of Our Changing Society. New York: Harper & Row.

De Praetere, T., Lobelle, C. (2010). The Dokeos e-Learning Project Management Guide (pp. 2007–2010).

eCollege (2016) Recuperado de: <http://www.ecollege.com/index.php>

educativa. (2016). Recuperado el 27 de abril de 2016 de: <http://www.educativa.com/>

Ellis, R. A., Ginns, P., & Piggott, L. (2009). E-learning in Higher Education: Some Key Aspects and Their Relationship to Approaches to Study. Higher Education Research & Development, 28(3), 303. Recuperado de: <http://www.informaworld.com/10.1080/07294360902839909>.

Fayyad, U., Piatetsky-shapiro, G., Smyth, P. (1996). From Data Mining to Knowledge Discovery. 17(3), 37–54.

Fidelis. (2016). Fidelis Inc. Recuperado el 16 de mayo 2016 de:

<http://www.fideliseducation.com/>

Fishtree. (2016). Fishtree Inc. Recuperado el 27 de abril de 2016 de:

<https://www.fishtree.com/>

García, A. (2013). De dónde venimos y hacia dónde vamos en Educación a Distancia.

Obtenido de contextos universitarios mediados:

https://www.youtube.com/watch?v=J_A2wp7_5Z8

GISMO. (2016). Graphical Interactive Student Monitoring Tool for MOODLE.

Recuperado el 17 de mayo 2016 de: <http://gismo.sourceforge.net/index.html>

Gobert, J. D., Sao Pedro, M., Raziuddin, J., & Baker, R. S. (2013). From log files to assessment metrics: Measuring students' science inquiry skills using educational data mining. *Journal of the Learning Sciences*, 22(4), 521-563.

Gómez, C. X., & Menéndez, J. R. (2011). Cuaderno de Prácticas de Análisis de Datos con SPSS.

Graduate XXI (2016). Recuperado de: <http://futuroeducativo.com/las-10-plataformas-de-aprendizaje-adaptativo-que-viajan-al-futuro-de-la-educacion/>

Growth Engineering. (2016). Recuperado el 27 de abril de 2016 de:

<http://www.growthengineering.co.uk/>

Hernández, R., Fernández, C. y Baptista, P. (2006). Metodología de la investigación (4ª ed.). México: McGraw Hill.

Hernández, J., Ramírez, M.J., Ferri, C. (2004). *Introducción a la Minería de Datos*. Pearson Prentice Hall.

Hernández Orallo, J., & Ferri Ramírez, C. (2006). Práctica de Minería de Datos. Introducción al WEKA. Universitat Politècnica de València, Valencia.

- Hilbert, M. (2012). Hacia un marco conceptual para las TIC para el desarrollo: lecciones aprendidas del “cubo” Latinoamericano. *Information Technologies & International Development*, 8(4), 261–280.
- Hoffmaster, D. (2006). The E-Learning Evolution. In *E-Learning Concepts and Techniques*. Pennsylvania.
- Honeycutt, B. (2013). The Flipped Approach to a LearnerCentered Class. Magna Publications.
- Hung, J.-L., Hsu, Y.-C., & Rice, K. (2012). Integrating Data Mining in Program Evaluation of K-12 Online Education. *Educational Technology & Society*, 15 (3), 27–41.
- IBM Corp. Released (2013). IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.
- Iglesias, A., Martínez, P., Aler, R., & Fernández, F. (2009). Learning teaching strategies in an adaptive and intelligent educational system through reinforcement learning. *Applied Intelligence*, 31(1), 89-106.
- IMS Global Learning Consortium. (1997). Recuperado de: <http://www.imsglobal.org/>
- Itmazi, J. (2005). *Sistema flexible de gestión del E-Learning para soportar el aprendizaje en las universidades tradicionales y abiertas*. PhD Thesis. University of Granada, Spain
- Jereb, E., & Šmitek, B. (2006). Applying Multimedia Instruction in e-Learning. *Innovations in Education & Teaching International*, 43(1), 15-27

- Kapp, K. (2003). Five Technological Considerations When Choosing an E-Learning Solution. *eLearn Magazine*, 7.
- Kerlinger, F.N. (1969). Research in Education. In R. Ebel, V. Noll, & R. Bauer (Eds.), *Encyclopedia of Education* (4th ed., pp. 1127-1134). New York: Macmillan.
- Kiffmeyer, Michael. (2004, November 9). The Evolution of e-Learning. Recuperado de: <http://knowledgemanagement.ittoolbox.com/documents/popular-q-and-a/theevolution-of-elearning-2902>
- Kulik, J. (1994). Meta-analytic studies of findings on computer-based instruction. En E. Baker & H. O'Neil, echnology assessment in education and training. Hillsdale, New Jersey, USA: Lawrence Erlbaum.
- Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American statistical Association*, 62(318), 399-402.
- León, O. y Montero, I. (1997) *Diseño de investigaciones: Introducción a la lógica de la investigación en psicología y educación* (No. 303.44). McGraw-Hill Interamericana.
- López, N., Lugo, M., & Toranzo, L. (2014). Informe sobre tendencias sociales y educativas en América Latina, 2014: políticas TIC en los sistemas educativos de América Latina.
- Lugo, M. T., López, N., & Toranzo, L. (2014). Informe sobre tendencias sociales y educativas en América Latina, 2014: políticas TIC en los sistemas educativos de América Latina. Recuperado de: <http://unesdoc.unesco.org/images/0023/002300/230080s.pdf#xml=http://www.unes>

co.org/ulis/cgi-

bin/ulis.pl?database=&set=0056DC79F1_1_45&hits_rec=3&hits_lng=spa

Lynch, T., & Lynch, C. (2002). Web-Based Education. *The Innovation Journal: The Public Sector Innovation Journal*.

Magdin, M., & Turcáni, M. (2015). Personalization of Student in Course Management Systems on the Basis Using Method of Data Mining. *Turkish Online Journal of Educational Technology-TOJET*, 14(1), 58-67.

Marenzi, I. (2014). *Multiliteracies and E-learning 2.0*. Frankfurt am Main: Peter Lang AG.

Macqueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1(14), 281–297.

Mavrikis, M., Patricia Charlton, & Demetra Katsifli. (2016, 2 18). *The Potential of Learning Analytics and Big Data*. Recuperado de:
<http://www.ariadne.ac.uk/issue71/charlton-etal>

MEFI (2016). PIH-MEFI Módulo 5. Recuperado de:
http://mefi.sel.uady.mx/pluginfile.php/3553/mod_lab

Naidu, S. (2006). *E-Learning, A Guidebook of Principles, Procedures and Practices*. Melbourne: Commonwealth Educational Media Center for Asia (CEMCA).

Ortiz, L. F. (2007). Campus Virtual: la educación más allá del LMS. *RUSC. Universities and Knowledge Society Journal*, 4(1), 3.

- Pearson education Inc. (2016). eCollege. Recuperado el 11 de mayo de 2016 de:
<http://www.ecollege.com/index.php>
- Pérez López, C. (2005). *Métodos estadísticos avanzados con SPSS*. Thompson. Madrid.
- Romero, C., & Ventura, S. (2006). *Data Mining In E-Learning*. Editorial WIT Press.
- Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, 51(1), 368–384.
- Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12–27.
- Romero, C., Espejo, P. G., Zafra, A., Romero, J. R., & Ventura, S. (2010). Web usage mining for predicting final marks of students that use Moodle courses. *Computer Applications in Engineering Education*, 21(1), 135-146.
- Sakai (2016). Recuperado de: <https://www.sakaiproject.org/learning-management>
- Sánchez, J. (2009). Plataformas de Enseñanza Virtual para Entornos Educativos. *Pixel-Bit: Revista de Medios y Educación*, 217-233.
- Sangra, A., Vlachopoulos, D., & Cabrera, N. (2012). Building an Inclusive Definition of E-Learning: An Approach to the Conceptual Framework. *International Review of Research in Open and Distance Learning*, 145-159.
- Siemens, G. (2005). Connectivism: A Learning Theory for the Digital Age. *International Journal of Instructional Technology & Distance Learning*, 2(1).

- Suhrman, S., Tutut , H., Haruna , C., & Jasni , M. Z. (2014). Data Mining for Education Decision Support: A Review. *International Journal of Emerging Technologies in Learning*.
- Schacter, J. (1999) the Impact of Educational Technology on Student Achievement. What the Most Current Research Has to Say. Milken Exchange on Education Technology. Santa Monica CA.: Milken Family Foundation.
- Shibley, I. (2009). Online Classroom. Magna Publication.
- Skillsoft Company (2016). Recuperado el 27 de abril de 2016 de:
<http://www.sumtotalsystems.com/>
- Smart Parrow. (2016). Smart Parrow PTY LTD. Recuperado el 02 de mayo 2016 de:
<https://www.smartparrow.com/>
- Statpoint Technologies, Inc. (2014). STATGRAPHICS Centurion XVII version 17.0.16.
www.statgraphics.com
- Trandafilii, E., Allkoçi, A., Kajo, E., Xhuvani, A. (2012). Discovery and evaluation of student's profiles with machine learning. City.
- Toronto University. (2002). ATutor: Learning Management System. Recuperado el 17 de Agosto del 2015 de <http://atutor.ca/atutor/>
- UADY (2010). Plan de Desarrollo Institucional 2010-2020. Obtenido de Universidad Autónoma de Yucatán: <http://www.pdi.uady.mx/docs/pdi.pdf>
- UADY (2012). Modelo Educativo para la Formación Integral. Obtenido de Universidad Autónoma de Yucatán: http://www.dgda.uady.mx/media/docs/mefi_dgda.pdf

- UADY (2013). Plan de Desarrollo UADY Virtual 2013-2020. Obtenido de:
http://uadyvirtualcloud.uady.mx/pluginfile.php/142/mod_resource/content/2/Plan%20Desarrollo%20UADY%20Virtual%202013-2020.pdf
- UNESCO (2002). *Information and Communication Technologies in Teacher Education. a Planning Guide*. Paris: Unesco.
- University of Cologne. (2004). Didactical concept of ILIAS. Recuperado de
http://www.ilias.de/docu/goto_docu_lm_392.html
- Universidad de Dortmund. (2001). RapidMiner. Retrieved November 22, 2012, from
<http://rapid-i.com/content/view/181/190/>
- Valsamidis, S., Kontogiannis, S., Kazanidis, I., Theodosiou, T., & Karakos, A. (2012). A Clustering Methodology of Web Log Data for Learning Management Systems. *Educational Technology & Society*, 15 (2), 154–167.
- Wang, J. K. (2014). A Research of Information Management System Based on Data Mining. *Applied Mechanics and Materials*, 616-619.
- Wenglinisky, H. (1998). Does it compute? The relationship between educational technology and student achievement in mathematics.
- Yildiz, O., Bal, A., & Gulsecen, S. (2015). Statistical and Clustering Based Rules Extraction Approaches for Fuzzy Model to Estimate Academic Performance in Distance Education. *Eurasia Journal of Mathematics, Science & Technology Education*, 11(2), 391-404.
- You, J. W. (2015). Examining the Effect of Academic Procrastination on Achievement Using LMS Data in E-learning. *Educational Technology & Society*, 18 (3), 64–74.

Witten I.H., Frank E., Hall, M.A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques (Third Edition)*.

Zapata, A. (2013). *Modelo Híbrido de Recomendación de Objetos de Aprendizaje*. Tesis doctoral. Universidad de Castilla-La Mancha, España.